National Library of Canada
Acquisitions and Bibliographic Services
395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque nationale du Canada
Acquisitions et services bibliographiques
395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence

Our file Notre référence

The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

L'auteur conserve la propriété du droit d'auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

0-612-22603-4

Canada

University of Alberta

Library Release Form

Name of Author: Mark Hicks

Title of Thesis: Genetic Variation in Managed Stands of Lodgepole Pine (*Pinus contorta)*

Degree: Master of Science

Year this Degree Granted: 1997

Permission is hereby granted to the University of Alberta Library to reproduce single copies of this thesis and to lend or sell such copies for private, scholarly, or scientific research purposes only.

The author reserves all other publication and other rights in association with the copyright in the thesis, and except as hereinbefore provided, neither the thesis nor any substantial portion thereof may be printed or otherwise reproduced in any material form whatever without the author's prior written permission.

Mark Hicks
8509-106 St. Apt. 306
Edmonton, Alberta
Canada, T6E 4J8

Oct. 2/97

University of Alberta

Faculty of Graduate Studies and Research

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research for acceptance, a thesis entitled *Genetic Variation in Managed Stands of Lodgepole Pine (Pinus contorta)* by *Mark Hicks* in partial fulfillment of the requirements for the degree of Master of Science in Molecular Biology and Genetics.

Ross Hodgetts

Ellen Macdonald

Curtis Strobeck

Bruce Dancik

Date: Oct. 2/97

This thesis is dedicated to my parents for without their understanding and overwhelming support none of this would have been possible. Thanks, Mom and Dad!

Abstract

Current reforestation practices have the potential to cause a reduction in the amount of genetic variation present in populations of lodgepole pine (*Pinus contorta*). An analysis of the genetic variation and allele frequencies of lodgepole pine was carried out using a technique called RAPD-PCR. Three different populations were used in this study including old growth, planted and naturally regenerated stands. Initial experiments were conducted to develop a rapid DNA extraction method which relied on small amounts of tissue. A total of 200 RAPD primers were screened to produce ten suitable polymorphic markers. The results from our analysis suggests that there is no difference in expected heterozygosity values between each of the different stands. However, further analysis into allele frequency distribution revealed some significant differences in old growth vs. planted and old growth vs. naturally regenerated stands. A model is proposed to explain these differences and provide some additional experiments.

Acknowledgments

I would like to acknowledge the help of my lab mates who made many long nights seem shorter and more bearable. Special thanks goes out to Sandra and Denise for the many hours of expertise provided and the many interesting conversations we had over coffee in the morning. Thanks to Andrew, for the many hours of computer instruction and interesting conversations. Thanks to Weldwood of Canada Ltd. and Weyerhaeuser Canada Ltd. for all of the sample information. Additional thanks to Canadian Forestry Service, Natural Sciences and Engineering Research Council through a grant to Ross Hodgetts and Ellen Macdonald under the NSERC/Forestry Canada/Industry partnership program. Also I would like to thank the Alberta Forest Development Research Trust Fund, Foothills Model Forest, and the Challenge Grants in Biodiversity from the department of Biological Sciences. Thanks to Lee and Aubrey for all the help with field collections. Special thanks to many of my friends for putting up with me during the lengthy writing process and a very special thanks to Leanne for all her help and patience.

Table of Contents

CHAPTER 1 INTRODUCTION 1

BIODIVERSITY ISSUES 1
OBJECTIVE OF MY THESIS 2
LODGEPOLE PINE 3
ASPECTS OF FORESTRY 8
DNA MOLECULAR MARKERS 16
RFLPS 18
RAPDS 22
MICROSATELLITES 26
LIMITATIONS 27

CHAPTER 2 MATERIALS AND METHODS 29

TISSUE COLLECTIONS 29
DNA EXTRACTIONS 38
AMPLIFICATION OF LODGEPOLE PINE ACTIN GENE 40
RAPD MARKERS 41
VERTICAL AGAROSE GELS 41
SOFTWARE ANALYSIS OF BANDING PATTERNS 42
ASSESSMENT OF GENETIC VARIATION 45
MICROSATELLITES 46
RADIOACTIVE LABELING OF MICROSATELLITE PRIMERS 47
AMPLIFICATION OF MICROSATELLITE MARKERS 47
ANALYSIS AND SIZING OF MICROSATELLITE ALLELES 48

CHAPTER 3 RESULTS 49

DNA EXTRACTION DEVELOPMENT 49
SELECTION OF RAPD MARKERS 58
SAMPLE SIZE ESTIMATION 58
TOTAL DNA EXTRACTIONS COMPLETED FOR ANALYSIS 65
GENETIC DIVERSITY ESTIMATES FOR LODGEPOLE PINE 69
CHI-SQUARE ANALYSIS OF RAPD VALUES 69
MICROSATELLITE DEVELOPMENT 77

CHAPTER 4 DISCUSSION 88

DNA EXTRACTION DEVELOPMENT 88
INITIAL GENETIC DIVERSITY ESTIMATES USING 2PQ 89
ANALYSIS OF TREATMENT POPULATIONS USING ALLELE FREQUENCIES 90
1. Comparisons of allele frequency among populations within treatments 90
2. Paired population analysis 91
3. Analysis of allele frequency between unharvested and seedling populations 93
MICROSATELLITE DEVELOPMENT 94

FUTURE WORK 95
BIBLIOGRAPHY **98**
APPENDIX A **108**

List of Tables

CHAPTER ONE INTRODUCTION **1**

1.1 PROPERTIES OF DNA BASED SYSTEMS FOR GENERATING GENETIC MARKERS ... 20

CHAPTER TWO MATERIALS AND METHODS **29**

2.1 WELDWOOD'S HINTON FMA COLLECTION REGIME 33
2.1 WEYERHAEUSER'S GRANDE PRAIRIE COLLECTION REGIME 34
2.2 NURSERY SEEDLING COLLECTIONS 37

CHAPTER THREE RESULTS **49**

3.1 SEQUENCE OF RAPD PRIMERS SCREENED (#1-100) 60
3.1 SEQUENCE OF RAPD PRIMERS SCREENED (#101-200) 61
3.2 TOTAL DNA EXTRACTIONS COMPLETED 66
3.3 AVERAGE HETEROZYGOSITY ESTIMATES FOR EACH TREATMENT PER LOCUS 71
3.4 CHI SQUARE ANALYSIS OF LODGEPOLE PINE POPULATIONS WITHIN TREATMENTS 76
3.5 COMPARISON OF CHI SQUARE VALUES FOR PLANTED AND NATURALLY REGENERATED STANDS 76
3.6 CHI SQUARE ANALYSIS OF COMPARISONS BETWEEN UNHARVESTED AND SEEDLING POPULATIONS 81
3.7 MICROSATELLITE SCREENING RESULTS OF SHORT REPEAT REGIONS IN LODGEPOLE PINE 85

List of Figures

CHAPTER ONE INTRODUCTION **1**

1.1 Geographical Distribution of *Pinus contorta* 7
1.2 Process of Domestication of Many Plant Species 13

CHAPTER TWO MATERIALS AND METHODS **29**

2.1 Forest Management Agreements in Alberta 31
2.2 Gel Pro Analyzer 2.0 Analysis Screen 44

CHAPTER THREE RESULTS **49**

3.1 Analysis of Lodgepole Pine Genomic DNA 55
3.2 Analysis of DNA Purity by PCR 57
3.3 RAPD Marker Profile 63
3.3 RAPD Marker Profile 64
3.4 Microsatellite Marker Profile for LP 13.14 87

List of Graphs

CHAPTER THREE RESULTS **49**

3.1 Sample Size Estimation of H-1-U Population 68
3.2 Average Heterozygosity Estimate over Ten RAPD Loci 73
3.3 Chi square Analysis of Paired Planted and Naturally Regenerated Populations of Lodgepole Pine 79

List of Abbreviations

" = inch

°C = degrees Celsius

A = adenosine

α = alpha

bp = basepair

BSA = bovine serum albumin

C= cystenine

Ci = currie

cm = centimeter

CTAB = cetyltriethylammonium bromide

dATP = deoxyadenosine triphosphate

dCTP = deoxycytosine triphosphate

dGTP = deoxyguanosine triphosphate

dTTP = deoxytyrosine triphosphate

DNA = deoxyribonucleic acid

EDTA = ethyldiaminetetraacetic acid

F2 = second generation

FMA = forest management agreement

γ = gamma

g = gram

G= guanine

h = hour

ha = hectares

Ltd. = limited

m = meters

mm = millimeter

M = molar

mM = millimolar

mmol = millimole

min = minute

mL = milliliter

n = haploid genome

ng = microgram

^{32}P = radioactive phosphorous

PCR = polymerase chain reaction

pH = negative log of hydrogen ion concentration in a given solution

pmoles = picomoles

QTs = quantitative traits

RFLP = restriction fragment length polymorphism

RAPD = randomly amplified polymorphic DNA

rpm = revolutions per minute

^{35}S = radioactive sulphur

sec = second

T = tyrosine

TBE = Tris Boric acid EDTA

TIFF =

UBC = University of British Columbia

µg = microgram

µl = microlitre

µM = micromolar

UV = ultraviolet

V = volts

Ver. = version

Chapter One Introduction

Biodiversity Issues

Genetic resources are the among the plant's most valuable raw materials, they are vitally important for allowing plants to adapt to changes in their natural environment. With the ever increasing effect that humans have on the genetic structure of many plant and animal species, it becomes ever more important that we evaluate and understand the role of genetic diversity that is present in all forms of life.

Some of the world's plant species have either undergone domestication, like corn and barley, or like lodgepole pine and white spruce are just starting to feel the effects of domestication. Typically, the biggest effect associated with selective breeding is the loss of genetic variation, a component of biodiversity. Isozyme analysis in inbred crop plants such as barley (*Hordeum vulgare* L.) has detected a loss in the genetic variation during domestication from wild populations to cultivars (Brown and Clegg 1983). In fact this is very common in many of the domesticated crop species and may become an important factor to consider during the initial stages in the domestication of many of Canada's conifer species.

In many countries, like Canada, tree breeding programs have begun in an effort to select for desirable traits in conifer species. Some of these traits include rapid and straight growth, disease resistance and large girth. However, high levels of genetic diversity, which are present during the early stages of breeding, must remain

throughout the breeding program if opportunities for selection in the future are to be maximized and inbreeding minimized (Adams 1981). Initial work has begun in trying to address the issue of preservation of species' gene pools. Among other agencies in various parts of the world, the research branch of the British Columbia forest service is developing a program for preserving the genetic resources both in the wild (parks, reserves, and gene resource management areas) and through tree improvement efforts (gene archives and test plantations) (Forests 1993).

Objective of my thesis

Forest populations have evolved through centuries of natural selection to contain a mixture of trees that match local environmental conditions. High levels of genetic variation are important to safe-guard against natural and man-made factors, like disease and pollution, especially in long lived crops such as trees (Hamrick and Godt 1990). A recent publication stated that there needs to be more research to document changes in heterozygosity under various types of forest management practices. Very few studies exist on the impact of harvest methods on genetic diversity (Ledig 1992). The monitoring of genetic diversity in tree breeding and production populations is needed, not only for gauging the impacts of breeding and production activities on reforestation stock, but also for effective genetic conservation of existing natural populations (El-Kassaby and Namkoong 1993; Millar and Marshall 1991).

Some 244 million ha of Canada's forested lands are considered productive and, of this, 91% is publicly owned and thus, the responsibility of provincial and federal governments (Canada 1990). Some 4000 logging firms, 80 pulp or newsprint mills, and 1160 sawmills and planning mills provide 756,000 jobs for Canadians (Canada 1990). These forested areas represent not only a huge environmental concern but very significance economic concern to all Canadians. Therefore anything that could potentially compromise the well-being of this resource should be monitored for the benefit of all Canadians.

The goals of this thesis are to address whether or not present management practices are having any effect on the patterns of genetic variation within and among populations of lodgepole pine, *Pinus contorta* var. *latifolia*. I set out to utilize DNA based markers to quantify genetic variation in several gene pools from different seed sources in a commercial harvesting lease in the province of Alberta. This area includes natural (uncut), naturally regenerated and planted populations of lodgepole pine. Genetic variation estimates were also determined from nursery stock which is destined for re-plantation of logged areas in northern Alberta. I believe that the unharvested areas provide the optimum genetic makeup for adaptation to future changes and that the culling and inadvertent favouring of certain plant types in the nursery and through the collection of seed could result in the reduction of genetic variation (biodiversity) in successive generations.

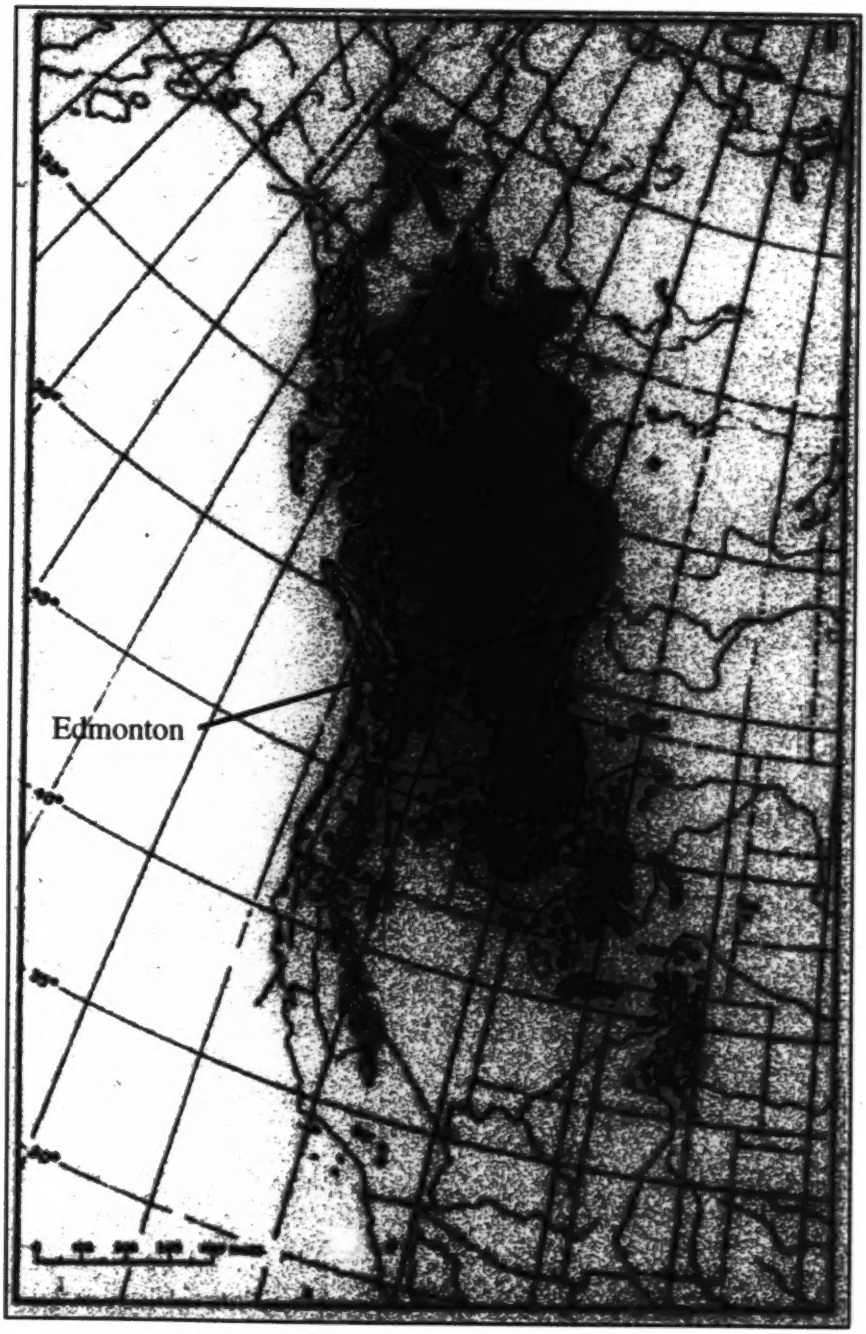
Lodgepole pine

The conifer species that was used in this study is one of the most extensively harvested forest trees in western Canada and can be recognized by its 1-3 inches long needles which are grouped in pairs. Its uses include interior paneling, exterior trim, particle board, posts, poles, framing material for light construction and it is used extensively for pulp and paper (Owens and Molder 1984). Lodgepole pine is considered a pioneer species because it is quick to occupy a site where natural or man-caused disturbances have created a condition of full sunlight and distributed seedbed. The species is characterized by its low tolerance of shade, rapid growth in young trees, slow growth in older trees and the ability to grow in almost any forest soil (Owens and Molder 1984). It produces large quantities of seed and good seed crops occur every one to three years. Lodgepole pine typically produces serotinous cones which require high temperature to release the seeds. This often follows fire leading to the rapid establishment of uniform, dense stands that often stagnate later when the trees are still quite small. The scales of the cones are stuck together firmly with resin often requiring heat (45-50°C) and drying before they will separate. The temperature at which the cones will open is highly variable and it is believed to be dependent on the degree of serotiny. This variation in serotiny has been shown to exist between individual trees within the same stand (Lotan 1975). Any particular stand may include a mix of serotinous trees and trees having cones that do open at maturity without the need for heat (Tinker *et al.* 1993).

Lodgepole pine (2n=24) is a member of the Pinaceae, the largest family of conifers and is one of the most familiar and important conifers in North America. It ranges from the Mackenize District of the Yukon in Canada, south to the southern Colorado and northern Baja, California Figure 1-1 (Owens and Molder 1984). It grows at a wide range of elevations from just above sea level to approx. 3400 m in the Rocky Mountains (Fowells 1965). Due to its wide distribution and differentiation, two to four geographical varieties or subspecies are commonly recognized (Pfister and Daubenmire 1975). *Pinus contorta* var. *contorta* or shore pine occurs along the Pacific coast and is characterized by its short and crooked stature with a flattened crown and non-serotinous cones. *Pinus contorta* var. *latifolia* to which the common name of lodgepole pine is generally applied is found throughout the interior regions and is tall with straight trunks and a narrow crown with serotinous cones usually. *Pinus contorta* var. *murrayana* are trees of the Cascade and Sierra Nevada mountains, and the *Pinus contorta* spp. *Bolanderi* from the Mendocino white plains in California may be considered separate interior varieties or subspecies, or may be combined with *Pinus contorta* var. *latifolia*. This distinction is based primarily on quantitative morphological traits but where the tree varieties co-exist they can and do interbreed freely.

Initial studies (El-Kassaby and Namkoong 1993) have looked at the genetic variation of managed forest tree species, like Douglas Fir, and seem to indicate that levels of genetic variation are very high for most species and that most of the variation is present within populations and a small amount exists among populations.

Figure 1-1. Geographical distribution of *Pinus contorta* in the western hemisphere. The shaded areas outline the areas with reported lodgepole pine growth. Adapted from (Owens and Molder 1984).


Edmonton

Other work has shown that conifers are considered to be one of the most genetically variable groups of species with at least 90 % of the genetic variation found within populations (Hamrick and Godt 1989). Its believed that the life history of long generations, large population sizes, high fecundity, long distance seed and pollen dispersal contribute to this pattern of variation (Hamrick and Godt 1989).

Aspects of forestry

The area of study for my thesis concentrates on what happens after the stand has been clear-cut. Clear-cutting is a method of tree harvesting where all trees of all species, irrespective of age and size are cut in one area and a new even-aged stand is allowed to develop. In the case of lodgepole pine, which is a pioneer species, and adapted to colonizing sites, this method of cutting is often quite successful. Once the area has been cut, regeneration of the trees can occur either artificially or naturally. Artificial methods (referred to as planted stands in this thesis) are preferred because they do not require the maintenance of a seed source for natural seeding when genetically improved species can be planted instead. This form of regeneration would allow new stands to be manipulated and created. Planting also has been shown (El-Kassaby and Namkoong 1993) to reduce regeneration delay because seedlings can be planted almost immediately following harvesting whereas if the area was left to regenerate naturally, the seeds would need to germinate first. However, studies show that this method of replanting can often lead towards a monoculture especially when only one species is planted (El-Kassaby and Namkoong 1993). Replanting may eliminate local patterns of variation, and may impact adjacent natural areas by pollen

and seed migration from non-local plantations (Ledig 1992). However, others (Bergmann and Ruetz 1991; Chaisurisri and El-Kassaby 1994; Muller-Strack 1987) feel that domesticated (improved) populations, while they will diverge from the wild-type, will be increasing the total diversity within the species because they will introduce rare alleles from other geographical locations. However, native vegetation may be displaced and the net result could be a loss in global diversity.

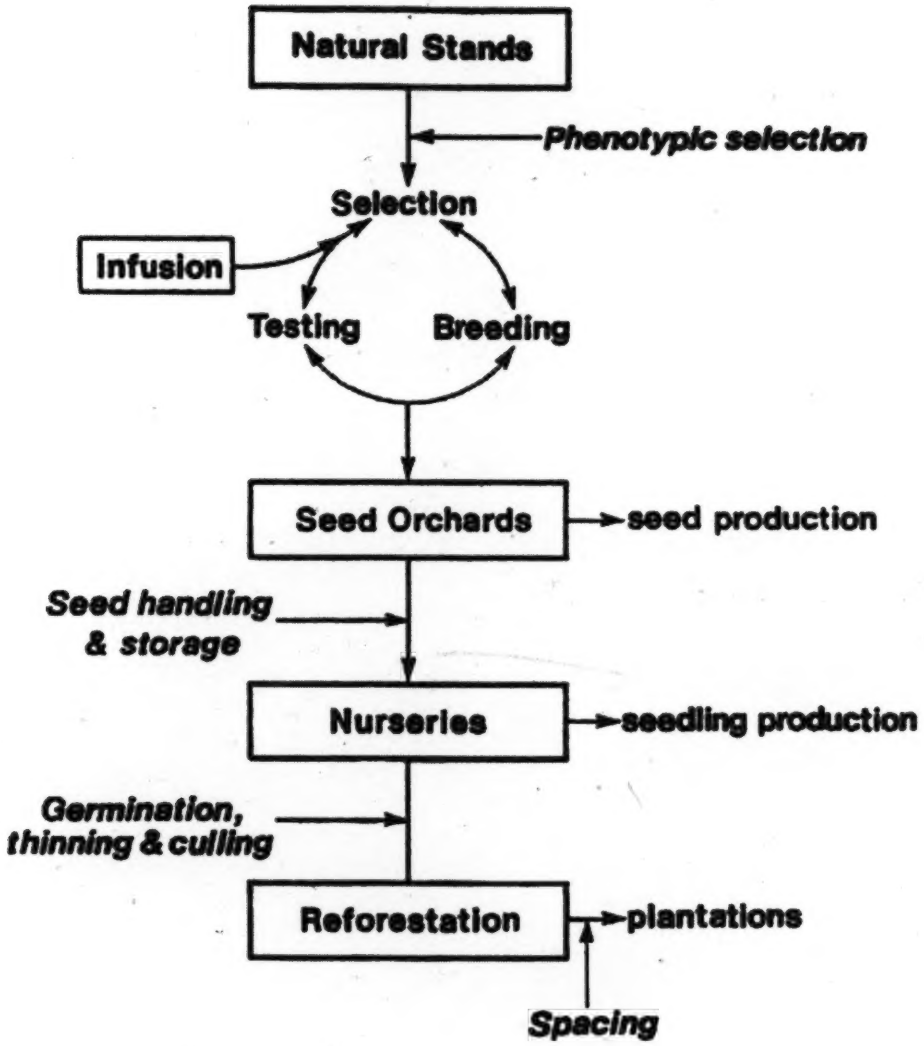
The other method of reforestation is through natural processes (referred to as naturally regenerated throughout this thesis) with some human intervention in the form of drag scarification which spreads the branches and cones all over the block. This method is dependent upon natural seeding of the clear cut area and is utilized for species with good seed dispersal capabilities. However, this method usually leads to irregular regeneration that is dense near the seed source and less dense as the distance from the seed source increases which results in regrowth that is not as uniform as with artificial methods (El-Kassaby and Namkoong 1993).

Both types of regeneration (artificial and natural) have the potential to alter the genetic make-up of present and future forests, especially with artificial regeneration, because the species and genotypic compositions can be manipulated and controlled. However, it is thought that the extent of change from stand to stand would be higher for artificial than natural regeneration (El-Kassaby and Namkoong 1993). In large clear-cut areas artificial regeneration will generally be done with some natural regeneration contributing to the new population from trees bordering the cut-block as

well as seed that was left on the ground. As a result the genetic structure of the new population will be a composite of local and other seed sources. Other problems are also associated with artificial regeneration. Studies have shown (Brown and Moran 1981) that planted populations are more susceptible to genetic loss than natural populations. Research done in Germany took beech seeds from local provenance trials and planted them in two different environments, in a greenhouse and under a natural beech stand, and after germination compared the isozyme markers from both seedling populations to the original seedlot. The seed populations that germinated in the woods were genetically different from those that were germinated in the greenhouse (Campbell and Sorensen 1984). The results suggested that significant selection between nursery-germinated plants and natural stands of beech had occurred. Yet, other research (Knowles 1985) portrays a different effect. Isozyme markers were used to determine if there are any differences between planted and natural stands of jack pine and black spruce. The results suggested that the genetic composition of both the plantation and natural stands is similar which indicates that the pressures that are believed to be present in a nursery situation, like time of germination, culling, spacing and transplantation trauma do not alter the genetic composition of the trees. Furthermore, this genetic similarity was observed in spite of the wide geographical make-up of the seed-zone collections which made up the plantation stock. However, the author is quick to point out that this analysis was done using only a few isozyme loci, seven for jack pine and five for black spruce and these results may be due to the fact this technique lacks the sensitivity needed to assess any differences between the two types of populations.

Current tree improvement methods could potentially reduce the genetic variation present depending on the breeding strategy and how this strategy is implemented (El-Kassaby 1992). The domestication of wild conifer species often follows several stages including phenotypic selection, a breeding program with its associated activities, seed and seedling production, reforestation and the establishment of plantations (Figure 1-2). Studies done (El-Kassaby, 1992) using western red cedar, Douglas-fir and Sitka spruce looked at each level of domestication to determine which, if any, stages have the most effect on genetic variation. Using isozyme markers El-Kassaby *et al.* were able to show that phenotypic selection was able to capture 90 % of the variation present in natural populations. The levels of variation in the selected populations were either higher, as in the case of sitka spruce, or similar, in western red cedar, when compared to natural populations. One specific area which was shown to have potential damaging effects on variation was in the seed-orchard. The authors suggests that random mating within the orchard must be maintained in order to mimic the natural process in the wild and try to prevent the loss of desirable genes. The process is often difficult to achieve because the clones must be sexually receptive at the same time and have similar reproductive output. This output must be equal for all parents or each clone will contribute differently to the seed crop and the amount of genetic change will be based on the extent of unequal cone production. In addition, due to the fact that most conifer species are wind pollinated, contamination from other areas is a major concern and must be controlled, along with self-fertilization and sib-matings. Because current forest practices are

Figure 1-2. Steps outlining the process of domestication of many tree species. Figure was adapted from (El-Kassaby 1992).


Natural Stands
Phenotypic selection
Selection
Infusion
Testing
Breeding
Seed Orchards
seed production
Seed handling
& storage
Nurseries
seedling production
Germination,
thinning & culling
Reforestation
plantations
Spacing

aimed at maximizing the economic value of the resource (El-Kassaby and Namkoong 1993), thinning is often used to remove low-breeding value clones from the orchard but in doing so it removes some genetic variation. Finally the storage of the seed can also affect variation. As in any storage situation there is some loss of vigor and if this loss is tree specific, one would expect additional loss in genetic variation during the storage of seeds (El-Kassaby and Namkoong 1993).

Probably the area with the greatest potential impact on variation is at the nursery during the germination and up to and including the culling stage. Many container nurseries rely on maximizing the total number of plantable seeds per unit area of the nursery such that any failure in germination in any container means a loss in productivity. As a result nurseries are placing two or more seeds into each growing cavity. This multiple sowing procedure ensures that there is at least one germinant present in all of the cavities. When two or more seeds are sown per cavity certain biological factors will come into play and may affect the genetic composition of the seedling crop. Differences in dormancy levels as well as germination speed will have dramatic effects on germinant size during thinning. This thinning, to leave only the largest seedlings behind, will result in unintentional selection for parents that produce a less-dormant, fast-germinating seeds, which may not necessarily produce the best tree (El-Kassaby 1996a). At the end of the nursery rotation, seedlings which do not meet "standards" are usually culled resulting in a further loss of parental genetic representation if the culled seedlings are from parents that are characterized as having slow early growth. Studies have been done which show that this can in fact happen. A

study (Piesch 1987) was done which looked at the height-performance among 45 Douglas fir families from the time of nursery lifting to age eight in the field. The results showed that some families that ranked lowest during the nursery phase ended up among the tallest in the field. In a recent study (El-Kassaby and Thomson 1990) it was demonstrated that there was a significant increase in selection pressure (bias to fast-germinating parents) when multiple sowing was performed. This was due to the nursery practice of culling small germinants such that fast-germinating seeds (which produce the biggest germinants at the time thinning) were favoured. Finally, some seedlings may not recover from transplantation shock as well as others, a stress to which they are never subjected naturally (Beineke and Perry 1966). Even if every seed or seedling was saved in the nursery, changes from the native pattern of genetic variation would still result because selective death occurs as a result of planting practices, and even the year of planting has been shown to have an effect. Groups of sugar maple have been shown to differ genetically and it is thought that this can be attributed to differences in selection resulting from year-to-year climatic fluctuations (Mulcahy 1975).

In addition to the nursery impact just described, the development of seed orchards to provide future seed stock can also have a drastic effect on the genetic composition of the seedlings planted (El-Kassaby and Namkoong 1993). Directional selection for growth and yield traits can potentially reduce the genetic diversity in seed orchard crops of domesticated species. However, it has been shown (Chaisurisri and El-Kassaby 1994) that phenotypic selection for quantitative traits (QTs) doesn't

provide a real threat to genetic diversity unless tree plantations are started from a very few select individuals. In fact, recent studies have revealed that levels of genetic diversity in production populations (i.e. plantations) are similar or higher than that present in natural populations (Chaisurisri and El-Kassaby 1994; Muller-Starck and Ziehe 1984). However, the authors from the 1994 paper are quick to point out that the seed orchard analyzed is still in its first stage of domestication and genetic diversity reduction might not be observed until after a complete cycle of recurrent selection (i.e. reduced number of parent trees in the 2nd generation seed orchard). This has been observed in other conifers like radiata pine (Moran and Bell 1987) and loblolly pine (Hamrick 1991). Additional problems can also occur in seed orchards including unintentional directional selection which can be caused by many factors including female fertility selection. This is known as the 20/80 rule, whereby, 20 % of the seed orchard clones provide 80% of the cone crop (Anonymous 1976). The end result is the reduction of the number of parents contributing to the seed crop thereby reducing the diversity present in a seed orchard.

DNA Molecular markers

Many types of marker systems have been used to assess or quantify the extent of genetic variation present in plant and animal species. Initially, phenotypic markers were used not only as a way to measure variation but also as a breeding tool. The use of classical phenotypic features for assessing variation or for following desirable traits has certain drawbacks. These markers can sometimes be influenced by environmental conditions. They require that the plant be grown to a suitable developmental stage

before certain characters can be scored and genetic control of many such characters is complex, such that their use in following traits in pedigrees is limited.

Initial efforts to find molecular genetic markers, relied on the polymorphic nature of proteins. These isozymes markers use neutral changes in the nucleotide sequence which alter the mobility of the enzyme in a starch gel without significantly altering the enzymatic activity of the protein. These resulting electrophoretic variants represent different allelic forms of that enzyme. While this early marker system allowed a glimpse at the variation present at the DNA level it did have a few short-comings. These markers only represent a portion of the genome i.e. the coding region and are thought to represent only enzymes involved with primary metabolism. While isozyme analysis is relatively fast and inexpensive, in some case, there are insufficient numbers of allozyme markers available to examine large portions of the genome (Liu and Furnier 1993). Furthermore, isozymes cannot detect all single base substitutions; about 23 percent of substitutions are synonymous and do not alter the amino acid because of the redundancy in the code and about 67 percent of amino acid replacements do not alter the net charge (Marshall and Brown 1975).

With many of these shortcomings a marker system that would allow direct assessment of the variation present at the DNA level was needed. DNA-based markers clearly allow the direct comparison of the genetic material of individual organisms avoiding any environmental influences on gene expression. A marker system was needed that would rely on DNA polymorphisms which occurred

anywhere in the genome including non-coding, single copy or repetitive DNA. The potential number of markers theoretically available is only limited by the amount of work, experimental skill and the money invested into finding them. In addition, polymorphic DNA based marker systems could provide direct experimental access to the region marked by them if the need ever arose (Bachmann 1994).

RFLPs

The first successful DNA-based marker system was reported in the early 1980s (Botstein *et al.* 1980) and was called Restriction Fragment Length Polymorphism of RFLP (Table 1-1). This type of marker is based on the variation in the position or restriction endonuclease sites among different genotypes or individuals. They are detected by hybridizing labeled DNA probes to Southern blots containing DNA digested with restriction enzymes. Identical-sized restriction fragments from different genotypes are interpreted as representing genetic similarities, whereas different-size fragments are interpreted as genetic differences. This procedure has numerous advantages in that the markers are co-dominant which means that heterozygotes can be distinguished from homozygotes and the system requires only 5-10 μg of DNA. Unlike phenotypic markers, RFLPs are unaffected by any form of environmental expression or suppression and allow a direct examination of the DNA and are fully penetrant. Yet, the system does have some drawbacks; there is little information on allele ranges and the distribution of the markers in the genome. It is assumed that the markers are distributed evenly over the entire genome but no direct evidence has been shown to prove this. There is also the problem of the need for

Table 1-1. Properties of DNA based marker systems. Adapted from (Rafalski and Tingey 1993).

Properties of DNA based systems for generating genetic markers

	RFLP	RAPD	Microsatellites
Principle	Endonuclease restriction Southern blotting Hybridization	DNA amplification with random primers	PCR of simple sequence repeats
Type of polymorphism	Alterations in restriction fragment	Presence or absence of amplified product	Stepwise changes in the length of the amplified product
Genomic abundance	High	Very High	Medium
Level of polymorphism	Medium	Medium	High
Dominance	Codominant	Dominant	Codominant
Amount of DNA required?	2-10 ug	10-25 ng	50 - 100 ng
Sequence information required?	No	No	Yes
Radioactive detection?	Yes/No	No	Yes/No
Development costs	Medium	Low	High
Start-up costs	Medium/High	Low	High

radioactive isotopes and the difficulty of automation (Rafalski and Tingey 1993). However, the biggest difference between isozymes and RFLPs is that each system detects genetic variation at different levels (Brown *et al.* 1989). Isozymes are gene products, while RFLPs reveal differences at the level of DNA in both coding and noncoding regions. Correlations between levels of isozyme and DNA polymorphisms have been reported (Chase *et al.* 1991) and (Messmer *et al.* 1991) but discrepancies between the two levels of polymorphisms have also been found (Brown *et al.* 1989). These results are none too surprising and indicate an important point which must be addressed when trying to compare genetic variation estimates from two very different marker systems. At the most they can be used to make only general statements about the variation present in any given organism.

Having said that, RFLPs have become a very important tool in forest genetics. They have been used to construct genetic maps in many conifer species including loblolly pine (Ahuja *et al.* 1994). In addition to map construction, the probes developed in loblolly pine have been used in other conifer species which allows for the direct comparison of genetic maps and other genetic information. They (Ahuja *et al.* 1994) were able to show that some of their probes could in fact hybridize to numerous conifer species indicating some form of gene conservation. RFLPs have also been used across a wide range of species to establish linkage relationships and to construct synteny maps (Ahn and Tanksley 1993).

RAPDs

The RFLP marker system, while effective, requires a large amount of time and is technically complex. What was needed was a marker system which didn't require radioactive isotopes and was technically simpler than RFLPs such that multidisciplinary groups could use this procedure to answer many population genetic questions. A procedure called Randomly Amplified Polymorphic DNA or RAPD (Welsh and McClelland 1990; Williams and et al. 1990) was developed which requires less time than RFLP assays and detects nucleotide sequence polymorphisms in a DNA amplification based assay. The system uses only a single primer of arbitrary nucleotide sequence and the amplification products can be separated on agarose gels stained with ethidium bromide. In the RAPD reaction, a single species of primer binds to the genomic DNA (15-25 ng) at two different sites on opposite strands of the DNA template. If these priming sites are within an amplifiable distance of each other, a discrete DNA product is produced (Tingey *et al.* 1992). This procedure relies on the statistical chance that the complementary primer sites occur in the genome as inverted repeats enclosing a relatively short stretch of DNA (Bachmann 1994). Using 10 base-pair primers is the most convenient since a total of 1,048,576 different arbitrary decamer primers can be made and finding a genetic polymorphisms in even the least variable population is only a question of persistence (Bachmann 1994). Using short primers and low annealing temperatures ensures that several sites, randomly distributed throughout the genome, give rise to amplification products. The polymorphisms that are present between individuals are the result of sequence differences in one or both of the primer binding sites, or an insertion or deletion

within the amplified region and are visible as the presence or absence of a particular RAPD band. Researchers (Williams *et al.* 1993; Williams and et al. 1990) have established that single base changes in RAPD primers results in different fragments being amplified, thereby providing direct support for the high stringency of the reaction conditions. However, RAPDs do have their problems. The presence or absence scoring method means that the technique provides only dominant markers, i.e., one cannot immediately deduce homozygosity or heterozygosity from the presence of a RAPD marker DNA band. This dominant nature of RAPD markers adds a factor of uncertainty to allele counts and therefore the presence of heterozygotes can not be directly determined from the agarose gel and must be inferred using acceptable population genetic equations like the Hardy-Weinberg equation.

RAPDs have many applications and have been used to create high-density genetic maps in many plant species (Reiter and Williams 1992). They have been used as molecular markers within regions containing genes of interest. This is made possible by the development of near-isogenic lines such that any RAPD marker that is polymorphic between two near-isogenic plants is potentially linked to the desired trait (Paran *et al.* 1991). This technique was used successfully to identify a RAPD marker linked to verticillium wilt resistance gene in tomato (Kawchuk *et al.* 1994). The difficulty in developing near-isogenic lines has been overcome by the use of RAPDs in a technique called bulk segregant analysis (Michelmore *et al.* 1992). This method involves screening for polymorphisms between two bulked DNA samples derived from a segregating population. Each bulk contains individuals selected to have

identical genotypes for a particular trait, e.g. disease resistance, but random genotypes at loci unlinked to the selected region. Therefore, the two bulked DNA samples differ genetically only in the selected region and are seemingly monomorphic for all other markers. Finally, the RAPD marker system has been used to look at genetic variation levels in numerous plant species, *Gliricidia sepium* and *G. maculata* leguminous trees native to Central America (Chalmers *et al.* 1992), snap bean (*Phaseolus vulgaris*) (Skroch and Nienhuis 1995) and buffalograss (Huff *et al.* 1993). RAPDs are markers powerful enough to differentiate individuals. A recent study successfully applied only two RAPD primers to detect sufficient genetic variation allowing for a complete differentiation between different pathotypes of *Peronospora parasitica*, a biotrophic fungal pathogen. (Tham *et al.* 1994) These results illustrate the potential value of RAPDs for detecting polymorphisms between organisms within the same species.

The application of RAPDs in many plant species extends to conifers and has helped to understand their complex genetics. The development of genetic linkage maps in conifers has been slow due to the large size of the genome, and the inherent difficulty involved in producing a segregating F2 population. However, researchers (Carlson *et al.* 1991) and (Chaparro *et al.* 1992) have shown that through the speed and efficiency of RAPD analysis it is possible to make mapping in conifers a reasonable endeavor. These markers have been used to asses the levels of genetic variation in red pine (Mosseler *et al.* 1992) as well. RAPD DNA fingerprint profiles have also been developed for white-Engelmann spruce species (Khasa and Dancik 1996). RAPDs have powerful applications in terms of DNA fingerprinting and clone

identification in numerous conifer breeding programs. RAPDs have been used to asses the genetic diversity of clones within Sitka spruce seed orchards and were able to distinguish each individual clone (Ven and McNicol 1995). This system has also been used to study the genetic variation in Norway Spruce (Bucci and Menozzi 1995).

Contamination is a big problem with RAPDs. The amplification products obtained with the RAPD method are from anonymous pieces of DNA, and since the decamer primers amplify DNA from virtually all sources which may include impurities such as bacteria and parasites in the material from which the DNA was isolated, the technique is very prone to contamination problems. The technique is also plagued with unreliable amplification products but if the conditions are standardized, and only loci that produce reliable products are used, many of the inconsistencies can be eliminated (Bachmann 1994). Difficulties with the assay reproducibility can be overcome by ensuring a consistency in DNA concentration, reaction conditions and thermal conditions during amplification. Even when banding patterns are reproducible there is no assurance of the identity of co-migrating amplification products in different samples. DNA can be isolated from a band, reamplified and labeled and hybridized to a Southern blot of the gel. It should label all the comigrating bands and usually does (Wilkie *et al.* 1993). However, there have been reports (Williams, Hanafey and Tingey 1993) where this is not true. A probe of the RAPD band was generated and a Southern was performed on the same gel and only some of the comigrating bands labelled with some different sized fragments also being lablelled.

In this case the net result is an under estimate of the extent of genetic variation present at that loci.

Microsatellites

Many of the problems associated with contamination, lack of co-dominant markers and reproducibility with RAPDs are avoided by the use of microsatellites markers. The first reported case where microsatellites were applied as PCR based DNA markers was in 1989, (Schlotterer and Tautz 1991), with a study of their inheritance in human pedigrees. Microsatellite repeats, also known as simple sequence repeats (SSRs), are usually composed of di- or tri-nucleotide repeats. They have been shown to be very abundant and highly polymorphic and are generally long tracts of the repeat motif but can be as short as two nucleotides in length. The basic unit is typically found to be repeated contiguously anywhere from 8 to 60 times, although these numbers can vary over even boarder ranges. It seems likely that this type of polymorphism is due to slippage mutations which are produced during the replication of DNA (Levinson and Gutman 1987). Other researchers believe this to be the case in the initial expansion of simple repeats but further expansion is caused by unequal crossing over (Hamada *et al.* 1984). As for a possible function of microsatellites they may be involved in gene regulation as they have been shown to enhance (Naylor and Clark 1990) or reduce (Pardue *et al.* 1987) the transcription of genes. They may also act as recombination hotspots (Gross and Garrard 1986), or play a role in the packaging and condensing of DNA into eukaryotic chromosomes (Wu and Tanksley 1993). Work done in rice (Terauchi and Konuma 1994) had shown that

the degree of polymorphism in microsatellites was significantly higher than in RFLPs such that a smaller set of markers can be used in mapping and/or linkage studies. The greater level of polymorphism in microsatellites compared to isozymes or RFLPs allows them to be used to potentially identify different forms of close kinship (parent-offspring, fullsib, half-sib) of individual clones in natural populations (Smith and Devey 1994). However acrylamide gels are needed for resolution of all of the alleles and large amounts of time, effort and expense are involved in identifying enough markers to do any sort of scientific analysis [Maughan, 1994 #82].

As for their use in conifers, not many microsatellite markers exist yet. Recent work has been done in the genus *Pinus* to look for microsatellites (Echt and May-Marquardt 1997). The researchers were quite successful in generating four markers however, there is a definite need for more of these markers in conifers. Additional work, (El-Kassaby 1996b), has been done in white pine to look at the frequency of different types of microsaatellites present in the genome. The results show that the di-nucleotide repeats $(AC)_n$ and $(AG)_n$ are the most commone with a predicted frequency of 2-4.5 sites/megabase pair (Mbp). With an estimated genome density for all microsatellites, excluding $(AT)_n$, at 16 sites/Mbp.

Limitations

With all of the advantages that exist with DNA markers, one major criticism remains. All of the DNA markers described here are considered neutral in that they don't have any direct connection to the traits (genes) of interest. However, if these

neutral markers are linked to selected loci then natural selection can change the frequency of essentially neutral markers and this is termed genetic "hitchhiking" (Mosseler, Egger and Hughes 1992). Direct evidence for "hitchhiking" does exist in the literature. A recent paper, (Mosseler, Egger and Hughes 1992) which looked at the levels of genetic variation present in red pine, reported extremely low levels of genetic variation present in RAPD markers. These results are none too surprising when you consider the fact that these plants may have undergone an extreme bottleneck (Mosseler, Egger and Hughes 1992). However, if these markers were indeed neutral and could undergo mutations without having any influence on surrounding genes, then one would expect to see higher levels of variation than were observed. The phenomenon of hitchhiking, the linkage of markers with genes, means that these markers are not neutral in terms of their ability to be good indicators of genetic variation.

RAPD markers, in addition to being considered neutral, have limitations with respect to their dominant nature. Therefore extreme caution should be applied when trying to assign an absolute genetic variation estimate to any organism based on RAPD data. Throughout this thesis the intention is to use the RAPD values as only relative numbers to be used for comparative purposes only and to be used between individuals which have been tested with the exact same markers.

Chapter Two Materials and Methods

Tissue collections

Lodgepole pine needle tissue was collected from two different areas, Weldwood of Canada Ltd.'s (Hinton) Forest Management Agreement (FMA) area and Weyerhaeuser Canada Ltd.'s (Grande Prairie) FMA located in the province of Alberta, Canada (Figure 2-1). A FMA is an agreement between the Alberta government and a timber company which allows the company to use forested land for the purpose of establishing, growing and harvesting timber. Three different types of stands or treatments were sampled: unharvested, planted and naturally regenerated. A stand is defined as a community of trees which share similarities with respect to composition, age, arrangement or condition such that they are distinguishable from the adjacent forest or other growth. Unharvested stands are defined as areas dominated by mature trees with an average mean age of at least one hundred years of age. Planted stands are defined as areas which were harvested (clear cut) then replanted with seedlings from a nursery while naturally regenerated regions are defined as areas which were first clearcut and then allowed to reseed via natural processes. Collections were set up such that planted and naturally regenerated stands were paired based on age, the area that was clearcut and geographical location. In each case paired treatments have the same number designation (Table 2-1).

Sample trees were chosen at regularly spaced intervals with a separation of at least 30 meters. A bearing was taken at the beginning of each stand collection to

Figure 2-1. Forest Management Agreements in the Province of Alberta, Canada. Blue-Weyerhaeuser Canada Ltd., Grande Prairie, Brown-Weldwood of Canada Ltd, Hinton

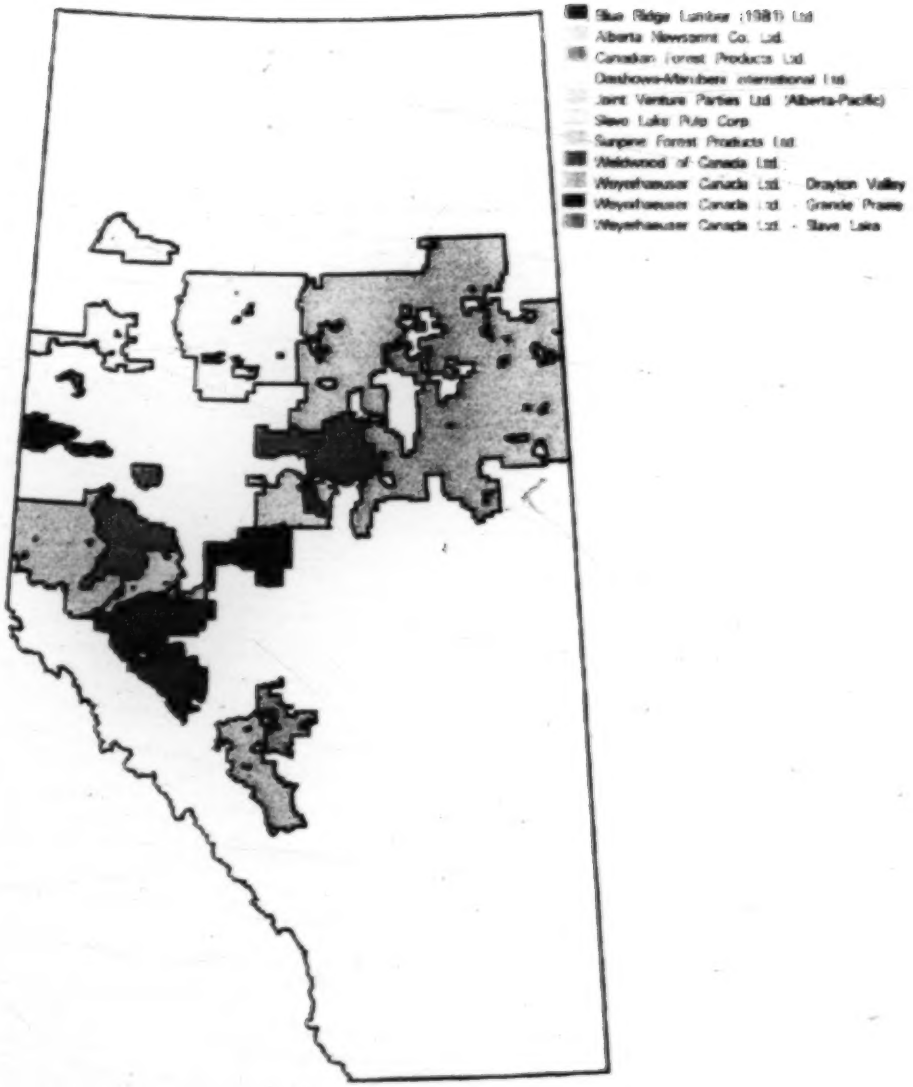

Blue Ridge Lumber (1981) Ltd
Alberta Newsprint Co. Ltd.
Canadian Forest Products Ltd.
Daishowa-Marubeni International Ltd.
Joint Venture Parties Ltd. (Alberta-Pacific)
Slave Lake Pulp Corp.
Sunpine Forest Products Ltd.
Weldwood of Canada Ltd.
Weyerhaeuser Canada Ltd. - Drayton Valley
Weyerhaeuser Canada Ltd. - Grande Prairie
Weyerhaeuser Canada Ltd. - Slave Lake

Table 2-1. Breakdown of the stand information and sample collection by FMA. (A)Weldwood of Canada FMA ,(B)Weyerhaeuser Grande Prairie FMA; N = number of samples collected per stand.T=township R=range

Weldwood's Hinton FMA Collection Regime

Location	Stand Designation	Date Harvested	Treatment Type	Treatment Date	Planting Date	Sample Number	N
T54 R25 Athabasca Working Circle Compartment 27	Old Growth	N/A	N/A	N/A	N/A	H-1-U	40
	Old Growth	N/A	N/A	N/A	N/A	H-2-U	40
	Old Growth	N/A	N/A	N/A	N/A	H-3-U	40
	Old Growth	N/A	N/A	N/A	N/A	H-4-U	40
	Old Growth	N/A	N/A	N/A	N/A	H-5-U	40
T54 R24 Athabasca Working Circle Compartment 16	Planted	1961	Scarified	1963	1970	H-1-P	40
	Planted	1966	Scarified	1967	1978	H-2-P	40
	Planted	1962	None	None	1969	H-3-P	40
	Planted	1965	Scarified	1977	1977	H-4-P	40
	Planted	1966	Scarified	1967	1974	H-5-P	40
T54 R24 Athabasca Working Circle Compartment 16	Naturally Regenerated	1961	None	N/A	N/A	H-1-R	40
	Naturally Regenerated	1969	Scarified	1970	N/A	H-2-R	40
	Naturally Regenerated	1962	Scarified	1964	N/A	H-3-R	40
	Naturally Regenerated	1966	None	N/A	N/A	H-4-R	40
	Naturally Regenerated	1965	Scarified	1966	N/A	H-5-R	40

Weyerhaeuser's Grande Prairie FMA Collection Regime

Location	Stand Designation	Date Harvested	Treatment Type	Treatment Date	Planting Date	Sample Number	N
T65 R08	Old Growth	N/A	N/A	N/A	N/A	G-1-U	40
T65 R08	Old Growth	N/A	N/A	N/A	N/A	G-2-U	40
T65 R08	Old Growth	N/A	N/A	N/A	N/A	G-3-U	40
T65 R07	Old Growth	N/A	N/A	N/A	N/A	G-4-U	40
T65 R08	Old Growth	N/A	N/A	N/A	N/A	G-5-U	40
T65 R06	Planted	1975	Site Prep	1977	1978	G-1-P	40
T65 R07	Planted	1977	Site Prep	1977	1979	G-2-P	40
T65 R07	Planted	1977	Site Prep	1977	1979	G-3-P	40
T65 R07	Planted	1978	Site Prep	1977	1979	G-4-P	40
T65 R08	Planted	1979	None	N/A	1980	G-5-P	40
T65 R06	Naturally Regenerated	1975	Scarified	1975	N/A	G-1-R	40
T65 R07	Naturally Regenerated	1975	Scarified	1975	N/A	G-2-R	40
T65 R07	Naturally Regenerated	1978	Scarified	1977	N/A	G-3-R	40
T65 R07	Naturally Regenerated	1978	Scarified	1977	N/A	G-4-R	40
T65 R08	Naturally Regenerated	1980	None	N/A	N/A	G-5-R	40

ensure that a straight line was followed to prevent any duplication of samples. A total of 40 samples was collected from each stand with five stands per treatment giving a total of 200 samples each from unharvested, planted and naturally regenerated areas. This process was repeated in the remaining FMA giving a total of 600 samples collected from each FMA.

A total of 1g of fresh needle tissue was collected from each tree. We tried to collect the most recent year's needles whenever possible. The tissue was then immediately packaged and placed on dry ice. Once the samples were returned to the lab they were then placed in a -70°C freezer until the DNA extractions could be carried out.

In addition to field collections, samples of seedlings grown in operational nursery conditions were also obtained from each company. Seedlings represented stock that would have been planted on the site where the field collections were made. Samples from Weldwood were collected from two different sources: one group of seedlings was produced at their local nursery located in the town of Hinton, Alberta and was collected in the spring of 1995 while the second set of samples came from a private nursery farm and was collected from cold storage in Edmonton, Alberta (Table 2-2). From these nursery collections, a total of 30 seedlings were taken at random from both Weldwood nurseries and used in this thesis.

Table 2-2. Samples collected from nursery populations from both Weldwood and Weyerhaeuser. Wild = seedlings produced directly from wild seed stock; improved = seedlings generated from seed produced from selected lines of trees.

Nursery Seedling Collections

Lumber Company	Name of Nursery	Sample Number	N	Type
Weldwood	Versa Cold	A	10	Wild
		B	10	Wild
		C	10	Wild
		D	10	Wild
		E	10	Wild
		F	10	Wild
		G	10	Wild
		H	10	Wild
Weldwood	Hinton Nursery	I	20	Wild
		J	20	Wild
		K	20	Wild
		L	20	Wild
Weyerhaeuser	Willow Valley	W-1	20	Wild
		W-2	20	Wild
		W-3	20	Wild
Weyerhaeuser	Willow Valley	W-ASOC-1	20	Improved
		W-ASOC-2	20	Improved
		W-ASOC-3	20	Improved
		W-ASOC-4	20	Improved
		W-ASOC-5	20	Improved
Weyerhaeuser	Beaver Lodge	G-1	20	Wild
		G-2	20	Wild
		G-3	20	Wild
		G-4	20	Wild
Weyerhaeuser	Beaver Lodge	G-ASOC-1	20	Improved
		G-ASOC-2	20	Improved
		G-ASOC-3	20	Improved
		G-ASOC-4	20	Improved

In addition to the field and nursery collections, ninety tissue samples were also obtained from the Huallen seed orchard site (G141 B1 P1) which is characterized as a "B1" low elevation lodgepole pine seed orchard. This orchard is made of half-sib families and includes 300 of the original 400 families selected on the basis of 5 year measurements of 4 progeny sites with a total of 15600 seedlings planted. Further thinning occurred and resulted in a total of 7800 trees remaining in the orchard. The tissue samples collected represent one tree from each of ninety different sites included in the original collections.

DNA extractions

Extraction of DNA from conifer pine needles is based on the method outlined by Mosseler (Hoopes and McClure 1981) with the following modifications. A total of 0.3g (fresh weight) of needles was diced into small pieces using a new razor blade. This tissue was then placed into 500 µl of ice cold grinding buffer (1 mM EDTA pH 8.0, 1.4 M NaCl, 100 mM Tris-Cl pH 8.0) in a 1.5 ml Eppendorf centrifuge tube. The tissue was homogenized for 3 seconds using a Janke & Kunkel "Ultra-Turrax" tissue homogenizer (model # TP10N) with a 8 mm diameter dispersing tool (model # 525N-8G) operating at a speed of 24,000 rpm. Following this, 100 µl of 10 M LiCl was added and the sample was homogenized for an additional 2 seconds at 24,000 rpm. Lysis was carried out by the addition of 500 µl of 60°C buffer (39 mM EDTA pH 8.0, 1.4 M NaCl, 100 mM Tris-Cl pH 8.0, 4% CTAB) and Proteinase K (Sigma) to a final concentration of 200 µg/ml and the sample was mixed by inversion. The mixture was

then incubated for 45 minutes at 60°C with frequent mild agitation. After incubation, 500 µl of chloroform-isoamyl alcohol (24:1) was added and mixed gently by inversion until the lower phase was green in colour. The emulsion was then centrifuged (14,000 rpm, 15 minutes, 25°C) and the aqueous phase was drawn off and re-extracted with 500 µl of chloroform-isoamyl alcohol. The aqueous phase was transferred to a fresh centrifuge tube and the nucleic acids were precipitated with 750 µl of ice-cold isopropanol and collected by centrifugation (6,500 rpm, 3 min, 25°C). The pellet was resuspended in 300 µl of TE buffer (10 mM Tris-Cl pH 8.0, 1 mM EDTA). RNase A was added to a final concentration of 10 µg/ml and the solution was incubated at 37°C for 20 minutes. The enzyme was extracted by the addition of 300 µl of chloroform-isoamyl alcohol (24:1) and after centrifugation (14,000 rpm, 5 minutes, 25°C), the aqueous phase was then transferred to a fresh 1.5ml centrifuge tube. The nucleic acids were reprecipitated by adding 50 µl of 7.5 M ammonium acetate, 1 ml of ice-cold ethanol and mixing gently. The pellet was collected by centrifugation (6,500 rpm, 3 minutes, 25°C) and resuspended in 300 µl of 200 mM ammonium acetate. A second ethanol precipitation was performed and the pellet was again collected by centrifugation (14,000 rpm, 5 minutes, 25°C) and drained. A 70% ethanol wash was performed and recentrifuged as before. The pellet was drained, dried under vacuum and resuspended in 50 µl of sterile double distilled water. The DNA was then used directly in RAPD-PCR reactions. If any reactions failed to amplify, a spermine precipitation was performed (Kenny 1986).

Amplification of lodgepole pine actin gene

Initially, DNA extracted from needle tissue was tested to determine its ability to be amplified before any RAPD analysis was performed. This test was done by determining whether or not the DNA could act as a template in amplification of the lodgepole pine actin gene. Primers were developed from the sequence of the 3' end of the lodgepole actin gene (Welsh and McClelland 1990; Williams and et al. 1990). The sequences of the primers that were used are (GHA 1) 5'-AGC CCT TGA CTA TGA GCA-3' and (GHA 2) 5'-AAC CTC CGA TCC AAA CAC-3' and they are expected to produce a product of 374 bp. Amplifications were carried out in 25 µl reactions containing 2 units of *Taq* polymerase, 50 mM Tris-Cl pH 8.0, 50 mM KCl, 1.5 mM $MgCl_2$, 0.01% Gelatin (w/v), 1 mg/ml BSA (Boehringer Mannheim), 2 µM each of dATP, dCTP, dGTP, dTTP (Pharmacia), 0.2 µM oligonucleotide primer and 200 ng of template DNA overlaid with a drop of heavy mineral oil (Fisher Scientific). Amplifications were performed in 0.5 ml PCR tubes (Rose Scientific) on a DNA thermal cycler (Grant Autogene II) using the following cycle parameters: an initial pre-soak of 94°C for 5 min followed by 94°C for 30 sec, 58.5°C for 45 sec, 72°C for 45 sec with a total of 32 cycles and a final soak at 72°C for 10 min.

The PCR product was then loaded onto a 1.2 % horizontal agarose gel and subjected to electrophoresis for 2 h at 3 V/cm. Gels were then stained in a dilute solution (0.5 µg/ml) of ethidium bromide for 45 minutes and photographed on a UV transilluminator using Polaroid 667 film.

RAPD markers

RAPD amplifications (Sarkar and Sommer 1990) were carried out in 25 µl reactions using the same PCR mixture and amount of template DNA that was used for the actin gene amplification. A control sample containing the reaction mixture without the template DNA was also included in each set of reactions. All of the reactions were set up in a laminar flowhood to minimize any possible chance of air-borne contamination. After using the flowhood it was sterilized using ultraviolet light (Maniatis *et al.* 1982). Amplifications were performed in a 96-well thin-wall polycarbonate PCR plate (Costar Thermowell G) sealed with high temperature sealing tape (Costar). The plate was placed in a DNA thermal cycler (Grant Autogene II) using the following cycle parameters: an initial pre-soak of 94°C for 4 min followed by 94°C for 30 sec, 39°C for 45 sec, 72°C for 2 min. for a total of 35 cycles and a final soak at 72°C for 10 min. The primers used throughout this study were purchased from the University of British Columbia and were originally developed by Dr. John Carlson at UBC.

Vertical agarose gels

The RAPD-PCR products were visualized on a vertical 1.5 mm thick 1.4% agarose (GIBCO/BRL/Life Technologies) gel and subjected to electrophoresis for 6 hr at 3 V/cm or overnight at 1.25 V/cm in TBE. The agarose gels were cast between two 5 mm thick plate glass plates separated by a 1.5 mm thick spacer. One plate measured

20 cm x 20 cm in size while the second measured 22 cm x 20 cm such that when the two plates were placed together a 2 cm overhang was formed at the top of the plate-gel-plate sandwich. To prevent slippage of the gel both plates were etched on the side which was in direct contact with the agarose gel. The plates were etched by blasting the surface with glass beads (#7 glass bead by Manus Co.) at 80 PSI. The wells were formed using a standard protein 15 well 1.5mm thick Teflon comb (Bio-Rad). A standard sequencing apparatus was used to run the gels (Tyler Research Instruments). After the Orange G loading buffer (Patel 1992) had just run off the bottom of the gel, the plates were removed from the apparatus, separated and the gel lifted off and placed into a dilute solution (0.5 µg/ml) of ethidium bromide for 30 minutes. The gel was then photographed using a digital imager ("The Imager" by Appligene, Oncor) and saved as a TIFF file for further analysis.

Software analysis of banding patterns

Analysis of the RAPD loci was performed using the Gel Pro Analyzer™ Ver. 2.0 software (Media Cybernetics). The program was used to identify PCR products on agarose gels and using a suitable size standard would assign a molecular weight to each band (Figure 2-2). Using these values, the software would group similar bands together. A default value of 3 percent was used throughout the project, such that if two bands differed by less than 3 percent, the software would consider both products to be from the same locus.

Figure 2-2. Gel Pro Analyzer 2.0 was used to determine sizes of each amplified product used in the experiment. The 100 bp size standard (Gibco/BRL/Life technologies) shown in lanes 1 and 12 is outlined in blue and the bands (lanes 2 - 11) for each of the samples are tagged with a red cross. The standard is then used to calculate the size for each of the tagged bands which are then binned together according to size ±3%. Arrows refer to polymorphic RAPD markers.

Assessment of genetic variation

As a measure of the genetic diversity of a population we computed its average heterozygosity using the RAPD data. Due to the fact that RAPD markers are not co-dominant, a direct measure of heterozygosity cannot be determined. It was estimated by scoring the absence of a band as a homozygous recessive individual and using the Hardy-Weinberg equation:

$$p^2 + 2pq + q^2 = 1 \qquad (1)$$

where p is the frequency of the dominant allele and q the frequency of the recessive allele, heterozygosity (2pq) could be calculated.

As another indication of genetic variation, the allele frequencies at the RAPD loci were compared between populations using a chi-square analysis at each locus and then, across all loci. I compared the frequency of the recessive allele (q) between populations using 2X2 or 2X3 chi-square tables.

The chi-squared test is used to ascertain the extent to which the observed numbers in a contingency table differ from expected numbers. Chi-square values can be calculated using the following formula:

$$X^2 = \sum \frac{(O-E)^2}{E}$$

Where "O" refers to the frequency of observed alleles and "E" refers to the frequency of expected alleles.

For the analysis of treatment homogeneity, one degree of freedom was used and a significance value of $P < 0.05$.

Microsatellites

A *Sau* 3A genomic library was constructed by Denise Adams using an M13 vector and clones were screened using two different probes: either a GT_n or GA_n probe. All positive clones were sequenced and analyzed to confirm the presence of microsatellite arrays (Denise Adams, pers. comm.) Once confirmed, primers were developed which flanked the locus using a software package (Primer Premier 3.1, Premier BioSoft) to obtain oligonucleotides of suitable length and base composition. The primers were generated and initially all of the forward primers were labeled. The potential microsatellite markers were then amplified on 10 random trees as well as the tree from which the library had been constructed. The products were run on a 6% arcylamide gel to determine if a product of the correct size was produced and if it was polymorphic.

Radioactive labeling of microsatellite primers

One of the pair of primers was 5' end-labeled at 37°C for 10 minutes, using T4 polynucleotide kinase (GIBCO/BRL/Life Technologies Cat. #18004-010). The reactions contained 50 pmoles of DNA and 2.5 µL of $\gamma[^{32}P]$ dATP (3000 Ci/mmol) (Amersham) in a total volume of 20 µL.

Amplification of microsatellite markers

PCR reactions were carried out in 10 µL reactions containing 2 units of *Taq* polymerase, 50mM Tris-Cl pH 8.0, 50 mM KCl, 1.5 mM $MgCl_2$, 0.01% Gelatin (w/v), 1 mg/ml BSA (Boehringer Mannheim), 2 µM each of dATP, dCTP, dGTP, dTTP (Pharmacia) and 40 ng of template DNA overlaid with one drop of light mineral oil (Fisher Scientific). In each case all reactions were assembled in a Nalgene LabTop Cooler™ Jr. to prevent any false priming from the *Taq* polymerase. First, the individual PCR tubes were pre-chilled in the cooler followed by the addition of template DNA. The chilled PCR mixture was then added and a drop of mineral oil was placed over the top of each reaction. Finally, the tubes were placed into a pre-heated (94°C) PCR machine (Grant Autogene II) and the amplification reactions were started. Once the cycling was completed, the reactions were stopped immediately by the addition of 10 µL of STOP buffer (0.3% Bromophenol Blue and Xylene Cyanol FF, 10mM EDTA (pH 7.5) and 97.5% deionized formamide) and stored at -20°C.

Analysis and sizing of microsatellite alleles

Each PCR reaction was subjected to electrophoresis on a 6% denaturing acrylamide gel containing 7M Urea using TBE at a constant power setting of 40W for 2.5 h. The gels were then dried and exposed to X-ray film (Fuji XAR) for 2 h. Sizing of each allele was done using a molecular weight ladder generated from a sequencing reaction of known DNA sequence. The template used was designated pDdc49b (Bendich *et al.* 1980; Murray and Thompson 1980; Taylor and Powell 1982). The sequencing reactions were carried out using the Pharmacia T7 Sequencing kit using 1 µL $\gamma[^{32}P]$ dATP (3000 Ci/mmol) (Amersham) instead of 1 µL of $\alpha[^{35}S]$ dATP. This change in isotope was done so that the exposure times of both the ladder and PCR products would be similar.

Chapter Three Results

DNA Extraction Development

An analysis of the genetic diversity in large numbers of individual trees relies on the need to isolate significant amounts of DNA both quickly and with minimal effort. When doing population studies using RAPDs, the most time consuming step is the isolation of DNA from numerous samples which, in my case, was on the order of 1200 samples. I set out to develop an extraction procedure that was inexpensive, quick and required only a few needles. To date previous procedures have relied on large amounts of foliage and were very time consuming, taking anywhere up to 3-4 days to complete.

Many current procedures are time consuming because they rely on an ultracentrifugation step in a CsCl gradient which takes 1 to 2 days (Dellaporta *et al.* 1983). Early protocols (Doyle and Doyle 1990; Murray and Thompson 1980) which did not require CsCl gradients used substantial amounts of tissue and 30 mL glass centrifuge tubes. Other procedures (Stewart and Via 1993), though simple, generally used large amounts of buffer (10 mL). A very fast and cheap protocol for isolating plant DNA is available (Murray and Thompson 1980), but it will only work on soft leaf tissue and not on needle tissue from gymnosperms like lodgepole pine. This is due to the fact that the method of tissue homogenization will not break up the needle tissue.

Plants have a cellular composition which is similar in many ways to animals but there are significant differences in a plant's chemical make-up which make it very difficult to isolate pure DNA. As a result many elaborate methods are currently used to remove inhibitors associated with the DNA after extraction. Some of these inhibitors are polysaccharides and tannins which pose a major problem in plants because they are difficult to separate from DNA (Do and Adams 1991). It has been shown (Hong *et al.* 1992) that the grinding of tissues releases many restriction enzyme and PCR inhibitors such as aromatic compounds, polyphenols, flavonoids and essential oils. They were able to show that neutral polysaccharides (dextran, gum sugar) were not very inhibitory when compared to acidic polysaccharides like pectin and dextran sulfate which were very inhibitory, even at low concentrations.

Due to the discovery of the inhibitors present in plant cells, numerous quick procedures were developed to circumvent their effects. LiCl had been shown to facilitate the softening of cell walls which has been shown to aid in the isolation of DNA from red seaweed and is believed to inhibit the association of polysaccharides with the DNA (Saunders 1993). Other procedures (Goodwin and Lee 1993) for removing polysaccharides involve a gel-purification step; however while inexpensive, this procedure is very time consuming because a gel must be run for each sample and then the DNA must be purified from the agarose gel. Another technique (Fang *et al.* 1992) includes a microwave step to alter the cell wall and membrane structure of the cell to aid in the recovery of DNA. However, I was were unable to obtain DNA that

could be amplified using this step. Additional procedures (Mosseler, Egger and Hughes 1992) utilize high-salt precipitation steps (1.0-2.5M NaCl) to remove polysaccharides and their inhibitory effects on restriction enzymes and *Taq* polymerase. The protocol we developed is outlined in detail in the Materials and Methods section of this thesis and incorporates several of the above procedures.

Essentially my procedure is based on a simple protocol (Murray and Thompson 1980) which is fast, since up to 40 samples can be processed from start to finish within an 8-hour day. The method is reliable since approximately 91% of the samples yield DNA that can be amplified by PCR. The protocol is easy to perform due to the fact that the entire procedure is carried out in microfuge tubes which, in addition to being fast, also diminishes the possibility for cross contamination between samples. The procedure works on a very small scale; one third of a gram of needle tissue is all that is required which allows for the non-destructive sampling of individual plants especially young seedlings. The technique uses CTAB (cetyltrimethylammonium bromide) which is a strong detergent believed to release DNA from membranes and proteins. Its use in plant DNA extractions was first reported by (Dellaporta, Wood and Hicks 1983; Doyle and Doyle 1990; Murray and Thompson 1980; Stewart and Via 1993) and allows the extraction of high molecular weight plant DNA without a costly and time consuming CsCl purification step.

Samples are initially homogenized using a tissue homogenizer designed to fit into a 1.5 mL microcentrifuge tube. The homogenizer is cleaned between each sample

using SDS and ethanol. This is the most time consuming step and potentially the most likely place for cross-contamination between different samples. The ideal situation would be to use disposable homogenizers but due to fiscal restraints this was not possible. Control PCR reactions were carried out between sample homogenizations to determine if any DNA was still on the homogenizer after cleaning. I failed to obtain any PCR products which indicates that the homogenizer was being cleaned sufficiently between consecutive DNA extractions.

Homogenization of the sample is the most critical step. Work was begun trying to use Mosseler's protocol because it used microcentrifuge tubes and was relatively cheap in terms of supplies needed. However, using the homogenization times stated in the protocol, we found that significant DNA shearing was occurring. It was later determined that this was due several factors, one of which was the presence of the CTAB detergent in the homogenization step. This resulted in the cell lysis and the release of large amounts of DNA into the solution during homogenization. Therefore, we decided to modify this step and removed CTAB from the homogenization step, introducing it into the grindate just before the incubation step (see Materials and Methods).

The second problem was the time and speed of homogenization. If the tissue is left in the homogenizer too long the DNA is sheared to an approximate size of 400 bp. However if the sample is not homogenized long enough, no DNA can be recovered. A time and speed for grinding was determined empirically based on the

genomic DNA's appearance on an agarose gel (Figure 3-1). A speed of 28,000 rpm for 3 seconds was determined to give the best results in terms of DNA integrity and amount.

Once a suitable homogenization time was achieved, the purity of the DNA in terms of its ability to be cut with restriction enzymes and its ability to act as a template in PCR was determined. Initial work in our laboratory (Gwenyth Harry, pers. comm.) done using early DNA extraction protocols (Hong *et al.* 1992) produced DNA that proved to be resistant to digestion and PCR amplification. The Mosseler protocol produced DNA of sufficient quality but it proved to be inconsistent. We decided to integrate some additional steps into the procedure in an effort to maintain DNA purity but more importantly to increase reproducibility. Work done in red algae (Kenny 1986) had shown that LiCl helped in preventing the association of polysaccharides with the DNA and as a result we decided to incorporate this into the 65°C incubation step. A few other simple modifications were also introduced including the addition of a RNase digestion step to remove any RNA present, as well as a 70% ethanol wash to further purify the DNA. The quality of the DNA was determined optically using a spectrophotometer and by its ability to be cut with restriction enzymes and to act as a template for PCR amplification. Initial DNA extractions performed using this new method were tested for digestion with *Eco* R1 and for PCR amplification of an actin gene. Earlier work on the 5' end of the lodgepole pine actin gene (Mosseler, Egger and Hughes 1992) had provided us with some genomic sequence to design the primers used to test the amplification ability of the DNA and a typical result is shown in Figure 3-2.

Figure 3-1. Analysis of lodgepole pine genomic DNA integrity after homogenization at 28 000 rpm for 3 seconds. Samples (Table 2-1, lanes 3-5) were run on a 0.8% TAE agarose gel with 1 kb ladder (GIBCO/BRL/Life technologies) (lane 2) followed by staining with ethidium bromide (0.5 µg/ml) for 20 minutes.

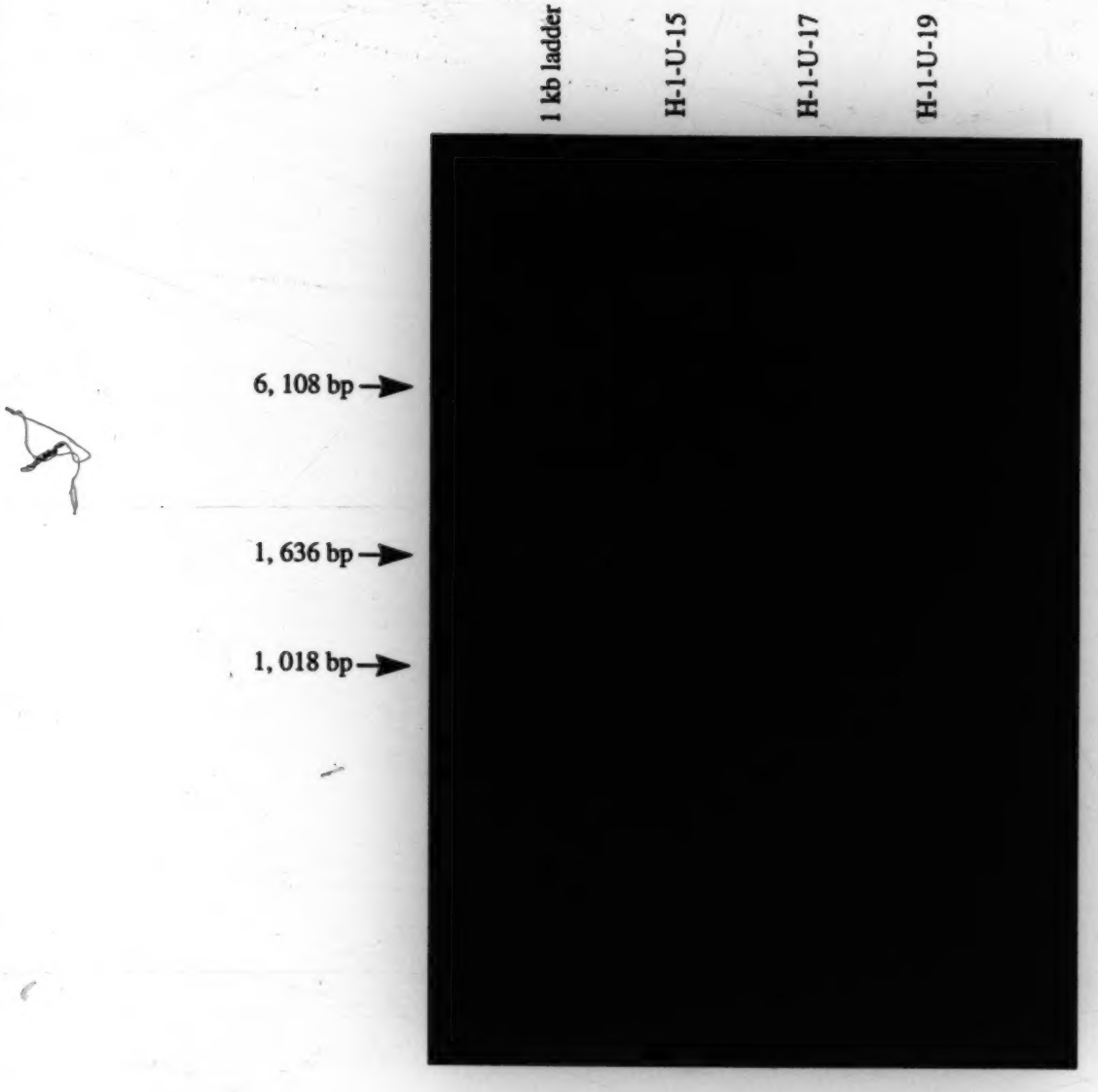

1 kb ladder
H-1-U-15
H-1-U-17
H-1-U-19
6, 108 bp
1, 636 bp
1, 018 bp

Figure 3-2. Analysis of DNA purity by its ability to support the amplification of the lodgepole pine actin gene. The resulting products were analyzed on a 1.4% TAE agarose gel stained with 0.5 µg/ml of ethidium bromide. Lanes 1 and 12, 100 bp ladder (GIBCO/BRL/Life technologies), lanes 2 through 9, samples from the old growth lodgepole pine population in Hinton, Alberta; lane 10, lodgepole pine DNA from genomic library; lane 11, no template DNA.

100 bp ladder
H-1-U-5
H-1-U-9
H-1-U-17
H-1-U-18
H-1-U-23
H-1-U-27
H-1-U-37
H-1-U-39
lp control
negative control
100 bp ladder
800 bp
400 bp
300 bp

The results from the PCR reactions and from the enzyme digests indicated that the extraction procedure met all of the criteria for which it was designed. It is fast, simple and inexpensive and in addition has been shown to work on different species including white spruce (data not shown), western hemlock and white pine (Darrin Gray, pers. comm.).

Selection of RAPD markers

An initial screening of three trees from the old growth population from Hinton, Alberta was used to determine which primers would produce the best results in terms of amplification, reproducibility and extent of polymorphism. A total of 200 primers was screened and a ranking system based on the method of (Biosoft 1995) was used to rate the RAPD products. The results revealed that 115 primers did not produce any discernible amplification products, 33 amplified poorly and the remaining 52 produced clear, intense bands following PCR amplification (Table 3-1). From this set of 52 primers an analysis was carried out to determine which primers produced reproducible polymorphic markers. Four were found which produced a total of 10 RAPD loci. The primers chosen for this study include primers #23, 63, 146 and 184 and their amplification products are shown in Figure 3-3.

Sample Size Estimation

Once suitable markers had been developed an initial analysis of one population from the old growth region in Hinton was completed to get an early estimate of the genetic diversity values for that given population. This was conducted

Table 3-1. Sequence of (5'-3') of DNA RAPD primers screened for production of RAPD markers (includes primers from 1 to 200.) N = no amplification, + = poor amplification (smearing with no finite bands), ++ = good amplification (distinct bands), +++ = excellent amplification (many distinct bands producing intense products). Primers highlighted in gray were used in this study.

Sequence (5'-3') of DNA primers screened for production of RAPD markers (primers 1-100)

Primer #	Sequence	Amp.	Primer #	Sequence	Amp.	Primer #	Sequence	Amp.
1	CCT GGG CTT C	+	35	CCG GGG TTA A	N	68	GAG CTC GCG A	N
2	CCT GGG CTT G	+	36	CCC CCC TTA G	N	69	GAG GGC AAG A	N
3	CCT GGG CTT A	++	37	CCG GGG TTT T	N	70	GGG CAC GCG A	+++
4	CCT GGG CTG G	N	38	CCG GGG AAA A	N	71	GAG GGC GAG G	+
5	CCT GGG TTC C	N	39	TTA ACC GGG C	N	72	GAG CAC GGG A	++
6	CCT GGG CCT A	+++	40	TTA CCT GGG C	N	73	GGG CAC GCG A	++
7	CCT GGG GGT T	N	41	TTA ACC GGG G	N	74	GAG CAC CTG A	N
8	CCT GGC GGT A	N	42	TTA ACC CGG C	N	75	GAG GTC CAG A	N
9	CCT GCG CTT A	+++	43	AAA ACC GGG C	+	76	GAG CAC CAG T	+
10	GGG GGG ATT A	N	44	TTA CCC CGG C	N	77	GAG CAC CAG G	+
11	CCC CCC TTT A	N	45	TTA ACC CCG G	N	78	GAG CAC TAG C	N
12	CCT GGG TCC A	++	46	TTA AGG GGG C	N	79	GAG CTC GTG T	N
13	CCT GGG TGG A	+++	47	TTC CCC AAG C	N	80	GTG CTC TAG A	N
14	CCT GGG TTT C	N	48	TTA ACG GGG A	N	81	GAG CAC GGG G	+
15	CCT GGG TTT G	+++	49	TTC CCC GAG C	N	82	GGG CCC GAG G	++
16	GGT GGC GGG A	++	50	TTC CCC GCG C	+++	83	GGG CTC GTG G	+
17	CCT GGG CCT C	+	51	CTA CCC GTG C	N	84	GGG CGC GAG T	++
18	GGG CCG TTT A	N	52	TTC CCG GAG C	N	85	GTG CTC GTG C	++
19	GCC CGG TTT A	N	53	CTC CCT GAG C	N	86	GGG GGG AAG G	+
20	TCC GGG TTT G	N	54	GTC CCA GAG C	N	87	GGG GGG AAG C	N
21	ACC GGG TTT C	N	55	TCC CTC GTG C	N	88	CGG GGG ATG G	N
22	CCC TTG GGG G	N	56	TGC CCC GAG C	N	89	GGG GGC TTG G	N
23	[illegible]	[illegible]	57	TTC CCC GAG G	N	90	GGG GGT TAG G	N
24	ACA GGG GTG A	N	58	TTC CCG GAG C	N	91	GGG TGG TTG C	N
25	ACA GGG CTC A	+	59	TTC CGG GTG C	+++	92	CCT GGG CTT T	+
26	TTT GGG CCC A	N	60	TTG GCC GAG C	N	93	GGG GGG AAA G	N
27	TTT GGG GGG A	N	61	TTC CCC GAC C	N	94	GGG GGG AAC C	++
28	CCG GCC TTA A	N	62	TTC CCC GTC G	N	95	GGG GGG TTG G	N
29	CCG GCC TTA C	N	63	[illegible]	[illegible]	96	GGC GGC ATG G	N
30	CCG GCC TTA G	N	64	GAG GGC GGG A	+++	97	ATC TGC GAG C	N
31	CCG GCC TTC C	+	65	AGG GGC GGG A	++	98	ATC CTG CCA G	N
32	GGG GCC TTA A	N	66	GAG GGC GTG A	N	99	ATC CCC TGG G	+
33	CCG GCT GGA A	N	67	GAG GGC GAG C	+	100	ATC GGG TCC G	N
34	CCG GCC CCA A	++						

Sequence (5'-3') of DNA primers screened for production of RAPD markers (primers 101-200)

Primer #	Sequence	Amp.	Primer #	Sequence	Amp.	Primer #	Sequence	Amp.
101	GCG GCT GGA G	N	135	AAG CTG CGA G	++	168	CTA GAT GTG C	++
102	GGT GGG GAC T	N	136	TAC GTC TTG C	N	169	ACG ACG TAG G	++
103	GTG ACG CCG C	+	137	GGT CTC TCC C	+	170	ATC TCT CCT G	N
104	GGG CAA TGA T	+	138	GCT TCC CCT T	N	171	TGA CCC CTC C	++
105	CTC GGG TGG G	+	139	CCC AAT CTT C	N	172	ACC GTC GTA G	N
106	CGT CTG CCC G	+	140	GTC GCA TTT C	N	173	CAG GCG GCG T	+++
107	CTG TCC CTT T	N	141	ATC CTG TTC G	N	174	AAC GGG CAG C	+
108	GTA TTG CCC T	N	142	ATC TGT TCG G	N	175	TGG TGC TGA T	N
109	TGT ACG TGA C	N	143	TCG CAG AAC G	N	176	CAA GGG AGG T	+++
110	TAG CCC GCT T	N	144	AGA GGG TTC T	N	177	TCA GGC AGT C	N
111	AGT AGA CGG G	+++	145	TGT CGG TTG C	++	178	CCG TCA TTG G	++
112	GCT TGT GAA C	N	[illegible]	[illegible]	[illegible]	179	TCA CTG TAC G	N
113	ATC CCA AGA G	N	147	GTG CGT CCT C	++	180	GGG CCA CGC T	++
114	TGA CCG AGA C	++	148	TGT CCA CCA G	N	181	ATG ACG ACG G	++
115	TTC CGC GGG C	+	149	AGC AGC GTG G	++	182	GTT CTC GTG T	N
116	TAC GAT GAC G	+	150	GAA GGC TCT G	++	183	CGT GAT TGC T	N
117	TTA GCG GTC T	N	151	GCT GTA GTG T	++	[illegible]	[illegible]	[illegible]
118	CCC GTT TTG T	N	152	CGC ACC GCA C	+	185	GTG TCT TCA C	N
119	ATT GGG CGA T	N	153	GAG TCA CGA G	+	186	GTG CGT CGC T	N
120	GAA TTT CCC C	N	154	TCC ATG CCG T	+	187	AAC GGG GGA G	+++
121	ATA CAG GGA G	N	155	CTG GCG GCT G	+++	188	GCT GGA CAT C	N
122	GTA GAC GAG C	++	156	GCC TGG TTG C	+	189	TGC TAG CCT C	++
123	GTC TTT CAG G	N	157	CGT GGG CAG G	N	190	AGA ATC CGC C	++
124	ACT CGA AGT C	+	158	TAG CCG TGG C	+	191	CGA TGG CTT T	N
125	GCG GTT GAG G	++	159	GAG CCC GTA G	N	192	GCA AGT CAC T	++
126	CTT TCG TGC T	N	160	CGA TTC AGA G	N	193	TGC TGG CTT T	N
127	ATC TGG CAG C	N	161	CGT TAT CTC G	N	194	AGG ACG TGC C	++
128	GCA TAT TCC G	N	162	AAC TTA CCG C	N	195	GAT CTC AGC G	N
129	GCG GTA TAG T	N	163	CCC CCC AGA T	+++	196	CTC CTC CCC C	++
130	GGT TAT CCT C	N	164	CCA AGA TGC T	N	197	TCC CCG TTC C	N
131	GAA ACA GCG T	+	165	GAA GGC ACT G	N	198	GCA GGA CTG C	+++
132	AGG GAT CTC C	+	166	ACT GCT ACA G	N	199	GCT CCC CCA C	+++
133	GGA AAC CTC T	++	167	CCA ATT CAC G	+	200	TCG GGA TAT G	N
134	AAC ACA CGA G	++						

Figure 3-3. RAPD marker profile produced as outlined in Materials and Methods and run on vertical 1.4% agarose gels followed by staining with 0.5 µg/ml of ethidium Bromide and captured by a video imager for later analysis. The 100 bp ladder(GIBCO/BRL/Life technologies) flanks either side of the gel to determine if there are any irregularities in the gel. The negative control is the same PCR reactions as other lanes in the gel except it contains no template DNA. Each RAPD gels was run with a negative control to ensure accurate band scoring. (A) primer #23, (B) primer #63, (C) primer #146, (D) primer # 184.

A)
100 bp ladder
negative control
H-2-U-23
H-2-U-21
H-2-U-20
H-2-U-17
H-2-U-16
H-2-U-15
H-2-U-14
H-2-U-8
H-2-U-3
H-2-U-2
100 bp ladder
1200 bp
1000 bp
800 bp
600 bp
L1
L2
L3
L4
B)
100 bp ladder
H-2-U-24
H-2-U-27
H-2-U-29
H-2-U-30
H-2-U-32
H-2-U-34
H-2-U-36
H-2-U-38
H-2-U-39
H-2-U-40
negative control
100 bp ladder
1200 bp
1000 bp
800 bp
600 bp
500 bp
L5
L6

C)
100 bp ladder
H-3-R-4
H-3-R-6
H-3-R-7
H-3-R-8
H-3-R-11
H-3-R-15
H-3-R-28
H-3-R-29
H-3-R-33
H-3-R-35
negative control
100 bp ladder
1200 bp
1000 bp
800 bp
600 bp
400 bp
L7
L8
D)
100 bp ladder
H-1-U-2
H-1-U-10
H-1-U-13
H-1-U-18
H-1-U-19
H-1-U-21
H-1-U-26
H-1-U-28
H-1-U-30
H-1-U-32
negative control
100 bp ladder
1200 bp
1000 bp
800 bp
600 bp
500 bp
L9
L10

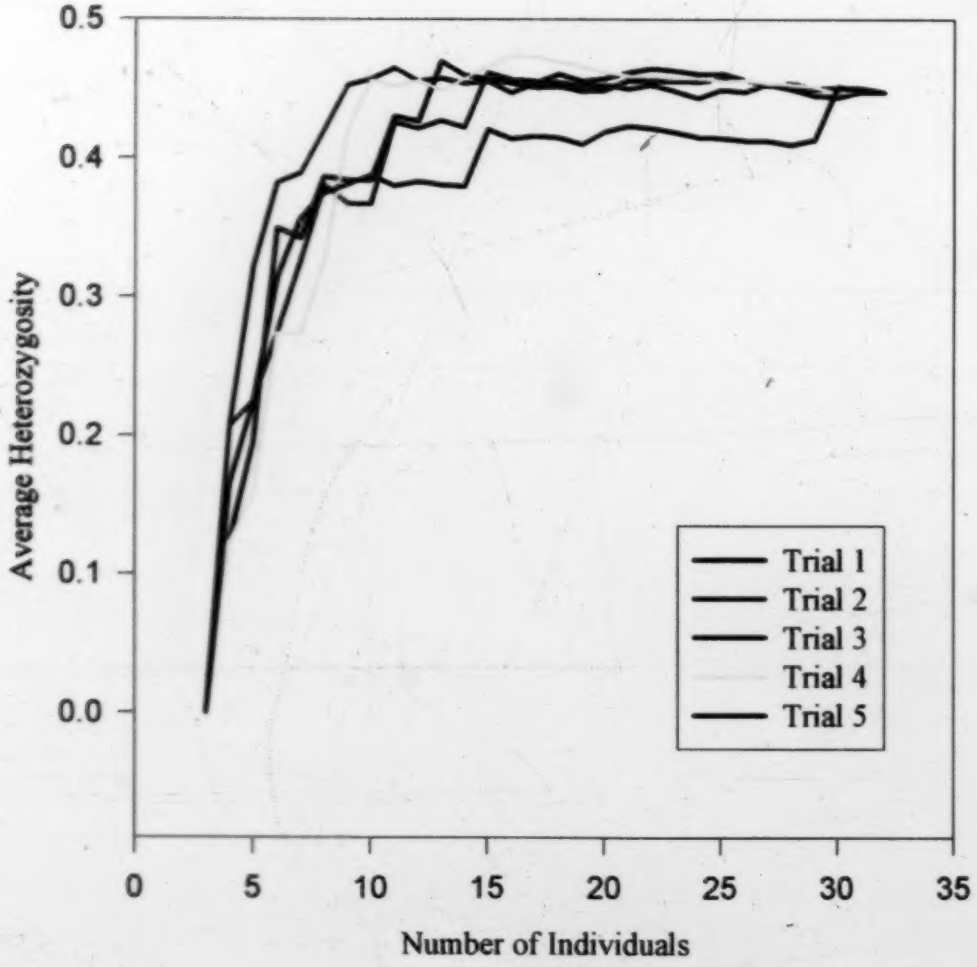
in an effort to determine the minimum number of individuals that were needed to represent all of the genetic variation present in a given stand. A bootstrapping method was employed which pulled samples at random from the stand and recalculated the genetic diversity as each sample was added. This process was repeated for a total of 5 trials and the results are plotted in Graph 3-1.

This analysis shows that a plateau in the average heterozygosity estimate for the population is reached by approximately the fifteenth individual. However, one of the five trials did not reach its plateau until the 29th individual was included, indicating a rare allele was not found until almost all of the samples had been analyzed. We were concerned about the results from the fifth trial and decided to be conservative and analyze thirty trees per stand to obtain a more precise estimate than using just fifteen trees would have produced.

Total DNA extractions completed for analysis

The results from the above analysis showed that the total number of DNA extractions that would need to be completed was considerably lower than the number of samples that had been collected. A breakdown of the 790 extractions is outlined in Table 3-2.

Graph 3-1. Average heterozygosity (genetic diversity estimate) of population H-1-U displayed as a function of the number of individuals. A total of five different sampling orders was used to generate the data. Each trial is designated by a different colour. The average heterozygosity of each tree is equal to the sum of the values over ten loci divided by the total number of loci.


0.5
0.4
0.3
0.2
0.1
0.0
Average Heterozygosity
0
5
10
15
20
25
30
35
Number of Individuals
Trial 1
Trial 2
Trial 3
Trial 4
Trial 5

Total DNA extractions completed

Treatment type	Population	No. of Samples Extracted
Hinton - Old Gowth	H-1-U	30
	H-2-U	30
	H-3-U	30
Hinton - Planted	H-1-P	30
	H-2-P	30
	H-3-P	30
Hinton - Nat. Regen.	H-1-R	30
	H-2-R	30
	H-3-R	30
Grande Prairie - Old Growth	G-1-U	30
	G-2-U	30
	G-3-U	30
Grande Prairie - Planted	G-1-P	30
	G-2-P	30
	G-3-P	30
Grande Prairie - Nat. Regen.	G-1-R	30
	G-2-R	30
	G-3-R	30
Hinton - Nursery	Total	80
Grande Prairie - Nursery	Total	80
Huallen Seed Orchard	Total	90
	Grand Total	790

Table 3-2 Breakdown in the number of DNA extractions completed per FMA.

Genetic Diversity Estimates for Lodgepole Pine

After all of the necessary extractions were completed, RAPD reactions were carried out on the Hinton samples only since time constraints precluded doing the Weyerhaeuser samples. The products were run on 1.4 % vertical agarose gels and the images captured by a digital camera for later analysis. At any given locus an individual was assigned either a one or a zero designation representing either the presence or the absence of a band respectively. As a result, a matrix profile for each sample was generated and a q^2 frequency could be calculated. Using the assumptions of Hardy-Weinberg where random mating is assumed to occur in an infinite population in the absence of any selective pressures, we calculated the expected heterozygosity values, or 2pq, for three Hinton populations and the seedlings from the Hinton nursery. The results in Table 3-3 and Graph 3-2 indicate that generally all of the loci showed similar values over the treatments except for locus six. The naturally regenerated stands show much lower heterozygosity when compared to any of the other treatments. An analysis of the individual populations within that treatment revealed that one population was composed solely of individuals that exhibited the PCR product at locus six, indicating that no homozygous recessive individuals were present.

Table 3-3. Average heterozygosity (genetic variation) estimates for each treatment per locus. L = locus, SD = standard deviation. A total of 90 individuals were analyzed in each of unharvested, planted and natrually regenerated while only thirty individuals were analyzed from the seedling population.

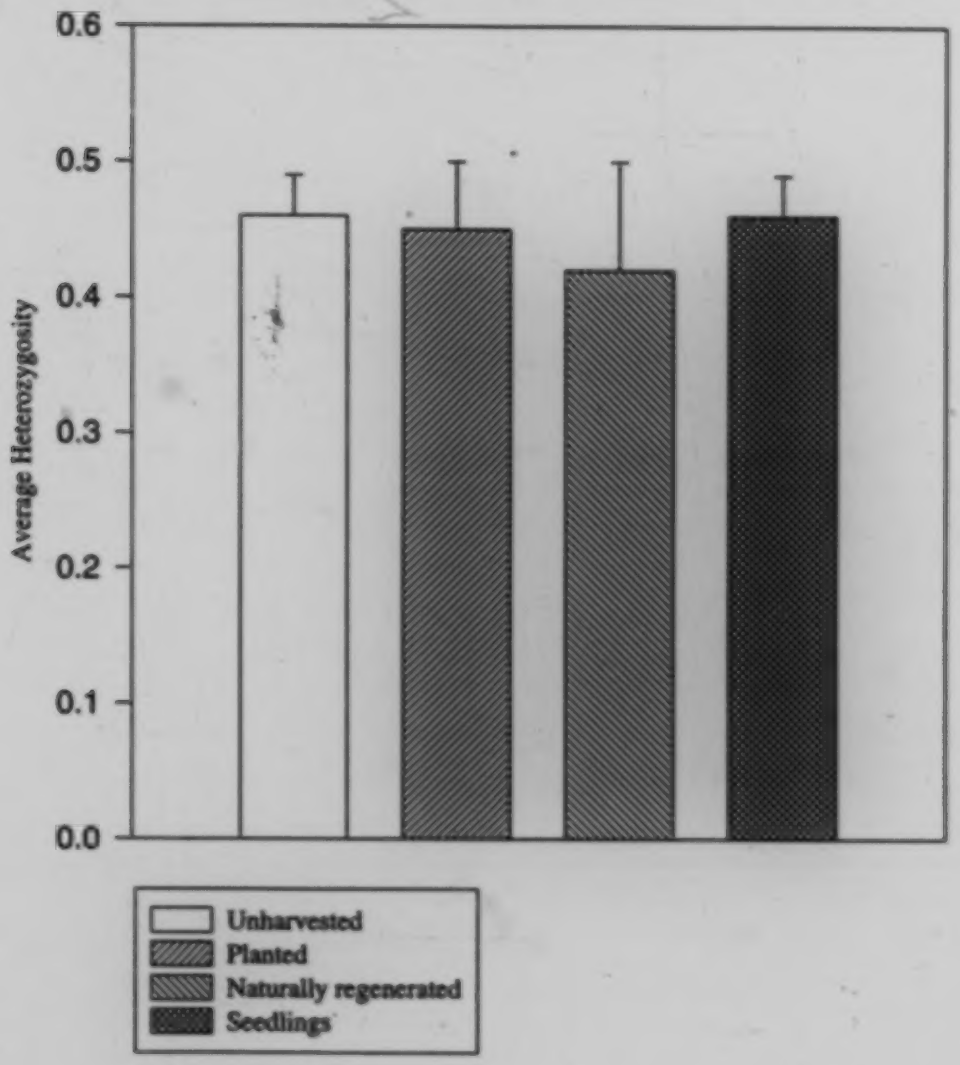
Average heterozygosity values for the Hinton population at each locus.

Population	L1	L2	L3	L4	L5	L6	L7	L8	L9	L10	Mean	SD
Unharvested	0.49	0.43	0.43	0.45	0.50	0.48	0.42	0.43	0.50	0.48	0.46	0.03
Planted	0.49	0.49	0.49	0.43	0.49	0.48	0.39	0.42	0.47	0.34	0.45	0.05
Natural Regen.	0.48	0.47	0.45	0.40	0.48	0.24	0.46	0.43	0.49	0.33	0.42	0.08
Seedlings	0.48	0.48	0.46	0.43	0.50	0.47	0.39	0.49	0.46	0.47	0.46	0.03
Mean	0.49	0.47	0.46	0.43	0.49	0.42	0.42	0.44	0.48	0.41		
SD	0.01	0.03	0.03	0.02	0.01	0.12	0.03	0.03	0.02	0.08		

Table 3-3 Average heterozygosity (genetic variation) estimates for each treatment per locus. L = locus, SD = standard deviation. A total of 90 individuals were analyzed in each of the unharvested, planted and naturally regenerated regions while only thirty individuals were analyzed from the seedling population.

Graph 3-2. Average heterozygosity estimate for each treatment type averaged over ten loci. Each estimate is based on a total of ninety individuals except for the seedling group which is based on thirty.

Average heterozygosity values for each treatment type averaged over ten loci.



Chi-square Analysis of RAPD values

In addition to the basic genetic variation estimates described above we decided to ask whether or not statistically different allele frequencies existed between treatments. Such differences might differentiate treatments that yielded identical genetic heterozygosity estimates. Using the matrix profile generated earlier, the frequency of homozygous recessive individuals (absence of a band) could be calculated. Since this value refers to q^2 we can calculate the observed frequency of the recessive allele by taking the square root of this value. Using the fact that $p + q = 1$ we can also calculate the frequency for the homozygous dominant allele (presence of a band). Any deviation from the expected number of alleles can be displayed as a chi-square value which can be tested to determine if it is significant. An example of the calculations carried out is given in Appendix A.

The first analysis was conducted to compare differences at each individual locu between the populations within each treatment using a 3x2 contingency table. In addition, a sum of the chi-square values over all ten loci was also obtained. The values calculated were determined to be significant at the 5 percent level. A value of 5.99 or greater is considered a significant difference with one degree of freedom. If the sum of the chi-squares is used, then the significance value is 18.31 or greater with ten degrees of freedom. The results are outlined in Table 3-4. These calculations were done to determine if treatment comparisons using pooled samples were possible. The results from Table 3-4 indicate that this is not possible because the differences between populations within each treatment are statistically significant.

Table 3-4. Chi-square analysis of lodgepole pine populations within treatments. Comparisons are made over 10 individual loci numbered L1 to L10. "SUM" refers to the sum of all of the chi-square values for the ten loci for each comparison. The shaded values refer to chi-square values that are significant. Data were generated from 90 individuals for all treatment types except for seedlings in which only 30 individuals were used.

Table 3-5. Chi-square values generated for paired planted and naturally regenerated lodgepole pine stands. A chi-square value for each locus was determined for each of the paired stand types. The shaded values refer to chi-square values that are significant. See Materials and Methods for pairing criteria. Data was generated from 90 individuals both treatment types.

Table 3-4 Chi-square Analysis of Populations Within Treatments.

Population Comparison	L1	L2	L3	L4	L5	L6	L7	L8	L9	L10	SUM
Unharvested	2.30	1.60	11.12	9.93	6.60	6.04	32.75	7.34	2.34	0.00	80.02
Planted	3.46	1.31	3.05	6.95	15.70	17.82	0.48	19.54	32.39	4.62	105.32
Nat. Regen.	0.24	13.00	0.65	4.13	2.37	12.96	27.85	21.26	4.95	7.27	94.18

Table 3-5 Comparison of Chi-square Values for Paired Planted and Naturally Regenerated Stands

Population Comparison	L1	L2	L3	L4	L5	L6	L7	L8	L9	L10	SUM
H-1-R vs. H-1-P	0.35	3.89	1.24	5.86	7.29	3.89	0.89	0.08	2.36	0.35	26.80
H-2-R vs. H-2-P	0.03	0.45	2.69	4.74	1.71	12.05	10.07	0.00	2.69	1.89	36.32
H-3-R vs. H-3-P	0.03	0.32	0.36	0.49	11.16	17.81	4.94	0.40	3.16	3.01	42.00

The second statistical analysis involved comparisons between paired populations of naturally regenerated and planted stands. Unfortunately comparisons with a paired unharvested stand were not possible, since these were no longer available in close geographical proximity to the regenerated stands. As Table 3-5 and Graph 3-3 show, the results from the comparisons among paired treatments would indicate that there are significant differences between paired stands (by age of stand and geographical location) that were regenerated by planting vs. natural regeneration. Thus the method of regeneration has a significant influence on the allele frequency of a population. However these differences are not global, only certain loci show significant differences while others, like L1 and L8, show almost no difference.

A final comparison was made between the seedlings and each of the three unharvested populations. The chi-square results are displayed in Table 3-6. These initial results seem to indicate that only three loci show significant differences in allele frequencies. These results contrast with the pairwise comparison between the planted and naturally regenerated (Table 3-5) which show significant differences at five loci. This may suggest that the seedlings are a relatively close representation of the allele frequency of the unharvested populations.

Microsatellite development

Denise Adams in our lab constructed a *Sau* 3A genomic library which was screened for dinucleotide repeats (Denise Adams, pers. comm.). A total of 40 positive

Graph 3-3. Chi-square analysis of paired planted and naturally regenerated populations of lodgepole pine. The red line indicates significance threshold. Anything below this line is considered non-significant and anything above this line is considered significant at the 5 percent level. Graph plotted using summed data from Table 3-5

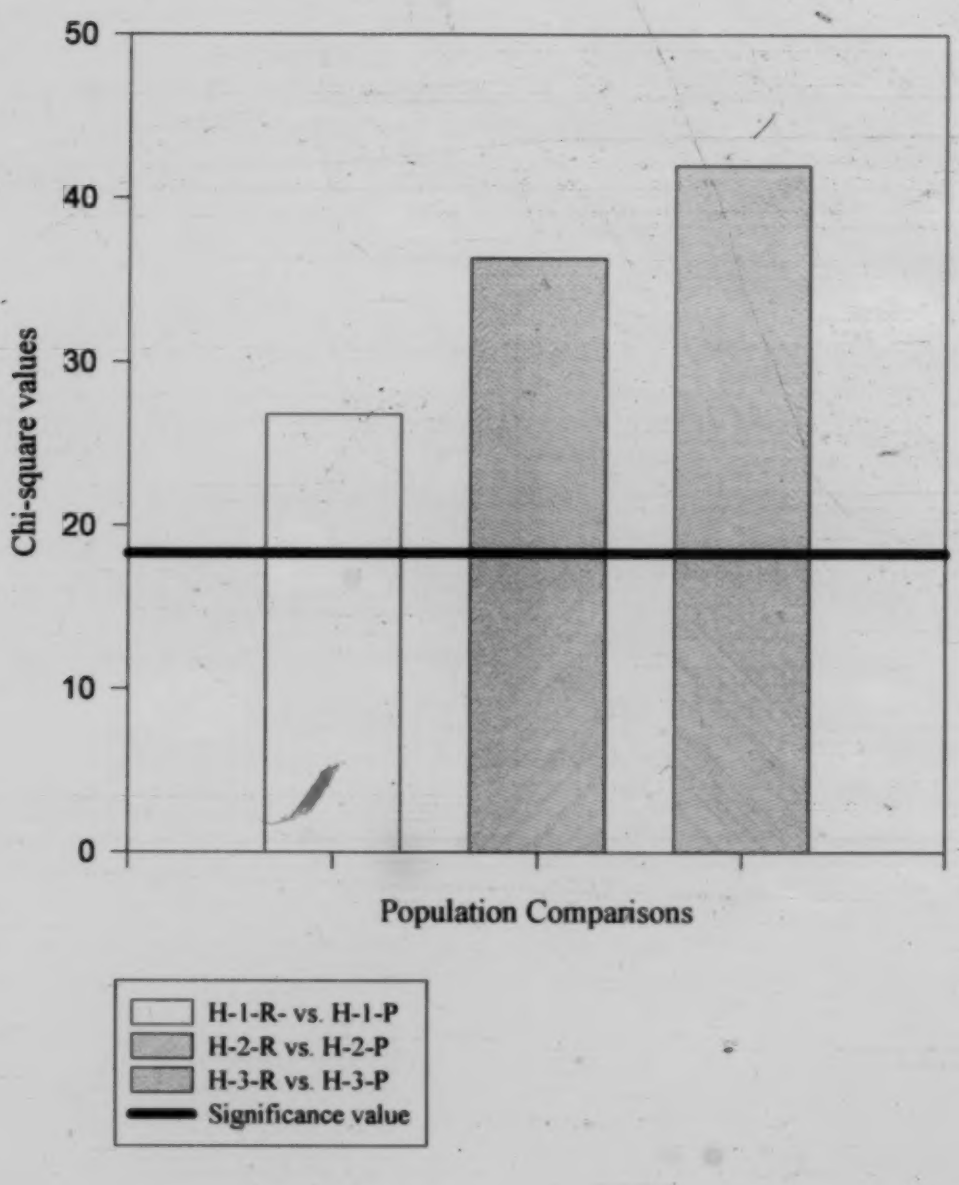

50
40
30
20
10
0
Chi-square values
Population Comparisons
H-1-R- vs. H-1-P
H-2-R vs. H-2-P
H-3-R vs. H-3-P
Significance value

Table 3-6. Chi-square values generated for population comparison between unharvested and seedlings. The shaded values refer to chi-square values that are significant. Data was generated from 30 individuals per population.

Table 3-6 Chi-square analysis of Unharvested vs. Seedling Populations

Population Comparison	L1	L2	L3	L4	L5	L6	L7	L8	L9	L10	SUM
H-1-U vs. Seedlings	0.31	3.26	3.41	0.24	3.75	0.71	4.38	5.85	1.11	0.01	23.03
H-2-U vs. Seedlings	0.31	0.31	1.94	1.96	0.26	0.71	0.00	0.00	6.53	0.01	12.03
H-3-U vs. Seedlings	0.54	1.24	0.54	3.03	1.77	1.75	17.53	0.73	3.07	0.01	30.21

phage clones were sequenced and analyzed for the presence of dinucleotide repeats. Denise began work on designing PCR primers for microsatellites which contained 20 or more repeats. Smaller repeats were left to be analyzed later if required. Denise ultimately was able to obtain only four markers and I decided to look at the smaller microsatellites in an effort to increase our marker base.

Primers which flanked the repeat regions in nine clones were designed using Primer Premier 3.1 (Tinker *et al.* 1993). All of the PCR reactions were carried out using the RAPD PCR mixture described in chapter two. This decision was made in an effort to standardize the PCR buffer being used thus making it easier and faster to set up numerous PCR reactions. All nine microsatellites were amplified and in each case we were able to produce a band of correct size from the lodgepole pine DNA from which the library was constructed (Table 3-7). These results indicate that the primers were correctly designed and were amplifying the correct regions. However, as the table indicates, I was only successful in generating one suitable microsatellite marker. The marker, LP 13.14 (DAD 15/16), produces a polymorphic band of predicted size based on sequence data from the clone (Figure 3-4). Furthermore the marker produces a simple profile with only one or two sutter bands which allows a more accurate and reliable allele identification. The LP 1.2 marker only produced a microsatellite stutter profile for the DNA from which the library was constructed. LP 5.6 was similar to marker LP 1.2 in that it only produced a band in the control lane and the stutter profile was very complex. Most microsatellite markers produced either one or two stutter bands which either proceed or trail behind the true band. Marker LP 5.6 produced

numerous stutter bands which proceeded and trailed behind what was believed to be the main band.

Table 3-7. Analysis of PCR products for potential microsatellite markers. Primers are listed in 5' to 3' direction with the forward primer listed first. Control = refers to the lodgepole pine from which the genomic library was constructed. Clone size = the size predicted from the sequence data for each microsatellite.

Table 3-7 Microsatellite Screening Results of Short Repeats Regions in lodgepole pine

Locus Name	Primers 5' - 3'	Repeat Size	Results
LP 1.2	lp 1 GAG CTA GCG AGA ACG GTC AA lp 2 CAT CGT GAC CAC CGG ATA AG	$(CT)_{17}$	stutter band produced on control only
LP 3.4	lp 3 TAT TGA CCC ACC AAC GTT GT lp 4 GTT GCA GGT TCA AAT TCA GC	$(CT)_{12}$	monomorphic band of predicted size
LP 5.6	lp 5 TCG GCC AAC AAA CTT ATG TC lp 6 AAA TCC AGA AAT CCG TAC CC	$(CT)_{9}$	non-characteristic microsatellite profile
LP 7.8	lp 7 CGG GGA GCT TCT GAA AAT AG lp 8 GCA TTT ACA CTG GGC ACC AT	$(CT)_{8}$	no PCR product except in control lane
LP 9.10	lp 9 GAA TAC AAC CAG CAT CCA GAC lp 10 AAG TGG CAA TAC TTG TGC TCA	$(CT)_{6}$	monomorphic band of predicted size
LP 11.12	lp 11 ATG CCT CCT CTA GCC CTT TA lp 12 CAT TGT GTA GAA GGT CGA CTG G	$(CT)_{11}$	monomorphic band of predicted size
LP 13.14 (DAD 15/16)	lp 13 AAT TCT TCA AAA GTT CAA T lp 14 TTA TGT TTC TGT TGT TTT A	$(CT)_{7}$	polymorphic band around predicted size
LP 15.16	lp 15 TCG CCC TCA ATG AAG TGT AT lp 16 AAG GTT GCC TCC AAG TTG TA	$(CT)_{9}$	monomorphic band of predicted size

Figure 3-4. Microsatellite marker profile for LP 13.14. PCR reactions were carried out using ^{32}P labeled primers and the products were run on 4% acrylamide gels. Autoradiography film was placed on the gel and exposed for two hours at room temperature. LP control refers to the lodgepole pine from which the genomic library was constructed. The sizing ladders were generated using the Pharmacia T7 sequencing kit using DNA of known sequence as a template, see Materials and Methods. The blank lane was used as a control to determine if there was any DNA contamination from other sources.

106 bp
104 bp
103 bp
102 bp
99 bp
98 bp
ladder ddATP
ladder ddCTP
LP control
H-1-U-1
H-1-U-3
H-1-U-5
H-1-U-12
H-1-U-14
H-1-U-15
H-1-U-18
H-1-U-23
H-1-U-32
H-1-U-39
negative control
ladder ddATP
ladder ddCTP

Chapter Four Discussion

The ever growing demand that we, as humans, place on our forests to provide us with many of the things that we use in our daily lives could potentially have far reaching effects on the genetic make-up of future forests. The demand for more wood products forces us to strive to produce trees which are bigger, stronger and healthier. The use of standard nursery practices has the potential to introduce strong selective pressures to select for trees with the highest potential economic return without any regard to the possible effects we may have on their complex genetic structure. Research needs to be done to address these potential problems to determine if they are real and, if so, how we can resolve them. This thesis was an attempt to evaluate some of these concerns and to provide some direction as to where additional research could be done in the future.

DNA Extraction Development

Probably the most technically challenging part of this thesis work was the development of a quick and simple DNA extraction protocol. The use of a tissue homogenizer and microcentrifuge tubes allowed us to quickly and easily extract DNA from numerous tissue samples. Without this procedure it would have taken considerably longer to isolate significant amounts of DNA suitable for PCR. Once the extractions were completed the RAPD reactions could be carried out and some initial estimates of genetic variation could be calculated.

Initial Genetic Diversity Estimates Using 2pq

To begin with, we looked at the expected heterozygosity values using the assumptions of Hardy-Weinberg (Table 3-3 and Graph 3-2). The results from each of the four different treatments suggested that the differences in the levels of variation were not large enough to indicate that current methods of forest management are having any effect on the levels of genetic variation. However, these estimates were calculated from the average of 10 different loci and as a result subtle differences in individual loci may not be obvious. If we take a closer look at each individual locus we see that nine out of the ten loci show very small standard deviation values which would suggest little difference in heterozygosity levels. There is, however, one locus (six) which shows a dramatic drop in genetic variation in the naturally regenerated stands. This was due to one particular population within the naturally regenerated treatment which failed to show any homozygous recessive individuals at locus six and is responsible for increasing the standard deviation value. The theory is that it was by chance that we failed to find any homozygous recessive trees with the current sample size and that if a larger sampling was taken it would be possible to obtain trees recessive at locus six. However, the frequency of the recessive allele will still be lower in that population.

Analysis of Treatment Populations Using Allele Frequencies

1. Comparison of Allele Frequencies Among Populations Within Treatments

Our initial comparison between treatments using 2pq to provide us with an estimate of genetic variation was flawed to a certain extent. While it is suitable to provide a rough estimate it can not detect changes in allele frequencies. For example, the most extreme case would be two populations with exactly the same expected heterozygosity but with completely opposite allele frequencies. Population one with p= 0.8 and q= 0.2 and population two with p= 0.2 and q=0.8. Both would give exactly the same heterozygosity value despite their allele frequencies being quite different. Our estimates of genetic diversity using the mean heterozygosity values in Table 3-3 revealed no changes. Therefore we decided that the best way to assess genetic differences between populations was by looking at allele frequencies.

Calculations in Table 3-4 revealed that it was not possible to group the samples from each treatment together and that any comparison between treatment types had to be done on a individual population basis. This is due to the fact that significant differences exist between each of the populations within each of the treatments which makes pooling populations and obtaining an average untenable.

2. Paired Population Analysis

A chi-square analysis was used to compare different paired populations from the planted and naturally regenerated treatments to ask if differences in allele frequencies are significant at the five percent level (Table 3-5 and Graph 3-2).

This comparison was made between paired populations based on geographical location and time of harvesting with the differences between H-3-R and H-3-P having the largest significant difference (42.00). Differences in allele frequency were seen in paired stands of different ages and different locations. However, this difference is not global, only certain loci show significant differences while others like L1 and L8 show almost no difference with levels of 0.03 to 0.35 and 0.00 to 0.40 respectively. Some of these differences could be explained by looking at where the seeds for each of these treatments come from. The naturally regenerated stands would come from surrounding populations of mature trees as well as cones present in slash from harvesting while the planted stands would represent a collection of artificial seed sources from nursery stock. However, these planted stands would also contain trees that arose from natural seed sources as well. This natural recruitment would be unavoidable and, like the naturally regenerated stands, would come from surrounding tree sources. This natural contribution may explain some of the similarities that are present between the different stand treatment types. An experiment could be designed to look at the amount of natural contamination in planted stands. The experiment would rely on the use of DNA markers to fingerprint each seedling before being planted in a given area. The location of each tree would be recorded such that in

successive years they could be located and re-fingerprinted. Any trees which did not have a known DNA fingerprint would represent natural recruitment and the extent of this influence from natural sources could then be determined.

Further work into understanding the selective pressures present in the naturally regenerated stands needs to be done to determine what is causing these changes in allele frequency. It is generally thought that the method of redevelopment of naturally regenerated areas relies on surrounding seed sources in addition to seed from trees cut down in the area. As was explained earlier, lodgepole pine is a pioneer species and therefore is well suited to the re-establishment of local populations. However, this species produces serotinous cones which require the presence of high heat, such as in forest fires, in order to melt the wax and allow the cones to open to release the seeds. Some research (Ledig 1992), was done on the heterogeneity present in the levels of serotiny in lodgepole pine cones and the results revealed high levels of variation in percent serotiny. If this is true for local populations of lodgepole pine around the Hinton area, then we can provide a possible model to explain the significant differences in allele frequencies present in naturally regenerated populations compared to other treatments. If there is indeed variation in the amount of serotiny, then following a clear-cut, trees with low levels of serotiny will be the most likely candidates to provide the seed for the following generation. In the absence of fire, those trees with the highest amounts of serotiny will not provide any seeds for the next generation. The result would be a change in allele frequencies in successive generations of lodgepole pine in that area. Obviously, the percent serotiny must be

determined in areas surrounding the clear-cut which is targeted for natural regeneration before this model would be considered as an explanation for the observed reduction in variation levels.

3. Analysis of Allele Frequencies between Unharvested and Seedling Populations

The hypothesis made at the beginning of this thesis was that current forest management practices are having some effect on the genetic variation of future populations of lodgepole pine. The effects of current nursery practices, which include multiple sowing, thinning and culling, on the levels of variation was not known and the research reported here was done in an effort to quantify the extent of these changes. The results from Table 3-6 suggest that current nursery practices are having only minor effects on the genetic composition naturally found in lodgepole pine. The results show that only three out of ten loci show significant differences when comparisons between the seedling population and the unharvested stands are made. This evidence strengthens the idea that in fact the seedlings currently being produced in Weldwood's nurseries have a similar allele frequency distribution as natural mature stands of lodgepole pine.

If we take a closer look at how the seeds for current nursery stocks are generated, we can provide a possible explanation for this observed similarity. Many nurseries use bulk seed sources which are composed of seed from numerous tree sources from similar elevations and geographic locations. As a result, the genetic

variation may in fact be higher than in natural populations because it may introduce some rare alleles. From a genetic variation point of view, this sounds good, but the use of bulk seedlots may in fact eliminate local patterns of variation, and may impact adjacent natural areas by pollen and seed migration from non-local plantations (Condit and Hubbell 1990). In fact, the local patterns of variation are obvious in the comparisons made between populations within each treatment (Table 3-4). It is clear that this species has a high degree of genetic diversity and strong patterns of local differentiation.

Microsatellite Development

The screening of libraries, by Denise Adams, with CA probes was done because recent evidence had shown that these types of microsatellites are common in tree genomes (Morchem *et al.* 1996). Their results are based on the screening of a number of tropical tree genomes with a range of 9×10^3 to 3×10^4 $(CA)_n$ sites per 5×10^9 bp of genomic DNA which is similar to the numbers found in humans and in beets (Echt and May-Marquardt 1997). Furthermore, recent work done, (Weber 1990), in white pine had shown that $(CA)_n$ and $(AG)_n$ motifs occurred at a frequency of 2-4.5 sites/megabase pair (Mbp). With the frequency of all microsatellites, minus $(AT)_n$ repeats, at 16 sites/Mbp. Microsatellites with larger repeat lengths were sought because it was thought that the larger repeats would show a higher number of alleles. Indeed it has been shown that for CA repeats the rate of strand slippage (repeat size increase) was found to increase rapidly when the number of repeats exceeded ten

(Weber 1990). These results would explain the data that I obtained with the relatively small microsatellites that I tested (Table 3-7). Half of the microsatellite repeats for which I made primers only produced a monomorphic band of the predicted size. However, this phenomenon can also be explained by the small nature of these repeats, it is quite possible that through an unequal crossing-over event the microsatellite could be lost and after PCR we would only see a monmorphic band. As for the primers which failed to produce a PCR product, the simplest explanation is of course is that the primers failed to either bind or amplify the region due to design problems. It is possible to redesign the primers and try again but due to time constraints, this was not done. Additonal work needs to be done using the microsatellite marker I developed in order to determine if the smaller repeat size has any effect on the number of alleles.

Future Work

The area which needs the most attention is to start the analysis of the Weyerhaeuser samples to determine if the trends seen in the allele frequency changes in the Hinton populations are also prevalent in the Weyerhaeuser samples. I think that the analysis should be done using microsatellites in both FMAs due to the many reasons outlined in the introduction to this thesis. Microsatellite are at this time the best DNA based marker system to answer the questions outlined in this thesis.

The results from the nursery samples, in which the allele frequency distribution was shown to be similar to unharvested regions, suggest that current

nursery practices are having no effect on allele frequencies. We suggest that in order to counteract these forces we should increase the variation present in nursery stock with the infusion of seed orchard, genetically improved and/or seed stock from a wider geographical area. Then we must look at each specific step in the nursery process and try to determine which methods to either eliminate or change to reduce their effects on the allele frequency distribution.

If the model that I propose, in which the heterogeneous mixture of serotinous cones in lodgepole pine contributes to the observed changes in allele frequencies in naturally regenerated stands, is to be verified it will require some additional work. An analysis needs to be done to determine the variation of serotiny in regions surrounding the naturally regenerated stands. If variation is seen it may provide further strength to the model and may question the use of this procedure in re-developing clear-cut areas.

Additional work could be done studying the composition of planted stands. After twenty years of growth it would be interesting to determine the number of trees within those stands which were planted and those which are from natural seed sources surrounding the stand. Genetic markers could be used to identify female trees in close proximity to the planted stands. These markers could then be used to trace the origin of all of the trees in the planted stand and determine which were planted and which are derived from local sources.

With the ever increasing demand that we place on our forests to provide us with the materials we need, the effects of harvesting on the genetic structure of many conifer species needs to be studied in detail. The early results from this study suggest that there may in fact be some problems with current practices and these areas will require additional study in order to determine the true extent of these problems. Fortunately we are at the early stages of domestication and if the current problems associated with many of our domesticated crops are to be avoided we must act quickly to preserve this resource for future generations.

Bibliography

Adams, W. T., 1981 Population genetic and gene conservation in Pacific Northwest conifers., pp. 410-415 in *Evolution today, Proceedings from the International Congression on Systematic Evolutionary Biology*.

Ahn, S., and S. Tanksley, 1993 Comparitive linkage maps of the rice and maize genomes. P.N.A.S **90:** 7980-7984.

Ahuja, M., M. Devey, A. Groover, K. Jermstad and D. Neale, 1994 Mapped DNA probes from loblolly pine can be used for restriction fragment length polymorphims mapping on other conifers. Theoretical and Applied Genetics **88:** 279-282.

Anonymous, 1976 Twenthieth annual report on the cooperative tree improvement and hardwood research program, pp. . North Carolina State University, Raleigh, North Carolina.

Bachmann, K., 1994 Tansley Review No. 63: Molecular markers in plant ecology. New Phytology **126:** 403-418.

Beineke, W. F., and T. O. Perry, 1966 Genetic variation in ability to withstand transplanting shock, pp. 106-109 in *Eighth Southern Conference on Forest Tree Improvement*, Savannah, GA.

Bendich, A. J., R. S. Anderson and B. L. Ward, 1980 Plant DNA: long, pure and simple, pp. 31-33 in *Genome Organization and Expression*, edited by C. J. Leaver. Plenum Press, New York.

Bergmann, F., and W. Ruetz, 1991 Isozyme genetic variation and heterozygosity in random tree samples and selected orchard clones from the same Norway spruce populations. Forest Ecology Management **46:** 39-47.

Biosoft, P., 1995 Primer Premier 3.1, pp. .

Botstein, D., R. White, M. Skolnick and R. Davis, 1980 Construction of genetic linkage map in man using restriction fragment length polymorphism. American Journal of Human Genetics **32:** 314-331.

Brown, A., M. Clegg, A. Kahler and B. Wier (Editors), 1989 *Molecular diversity in plant populations*. Sinauer Assoc, Sunderland, Mass.

Brown, A. H., and G. F. Moran, 1981 Isozymes and the genetic resources of forest trees, pp. 1-10 in *Isozymes of North American forest tress and forest insects*, edited by M. T. Conkle. Pacfic Southwest Forest Experimental station.

Brown, A. H. D., and M. T. Clegg, 1983 Isozyme assessment of plant genetic resources, pp. 285-295 in *Isozymes: Current Topics in Biology and Medical Research*, edited by A. R. Liss, New York.

Bucci, G., and P. Menozzi, 1995 Genetic variation of RAPD markers in a *Picea abies* Karst. population. Heredity **75:** 188-197.

Campbell, R. K., and F. C. Sorensen, 1984 Genetic implication of nursery practices, pp. 183-191 in *Forestry Manual: Production of Bareroot Seedlings*, edited by M. L. Duryea, and D. L. Thomas. Martinus Nijhoff/Cr. W. Junk Pulishers.

Canada, F., 1990 *Forestry Facts*. Supply and Services Canada.

Carlson, J., L. Tulsieram, J. Glaubitz, V. Luk and C. Kauffeldt, 1991 Segregation of random amplified DNA markers in F1 progeny of conifers. Theoretical and Applied Genetics **83:** 194-200.

Chaisurisri, K., and Y. A. El-Kassaby, 1994 Genetic diversity in a seed production population vs. natural population of Sitka Spruce. Biodiversity and Conservation **3:** 512-523.

Chalmers, K., R. Waugh, J. Sprent, A. Simons and W. Powell, 1992 Detection of genetic variation between and within populations of *Gliricidia sepium* and *G. maculata* using RAPD markers. Heredity **69:** 465-472.

Chaparro, J., P. Wilcox, D. Grattapaglia, D. O'Malley, S. McCord *et al.*, 1992 Genetic mapping of pine using RAPD markers: Construction of 191 marker map and development of half-sib genetic analysis., pp. in *Advances in gene technology: Feeding the world in the 21st century. Miami Winter Symposium*, Miami, Florida.

Chase, C. D., V. M. Ortega and C. E. Vallejos, 1991 DNA restriction fragment length polymorphisms correlate with isozyme diversity in *Phaseolus vulgaris* L. Theoretical and Applied Genetics **81:** 806-811.

Condit, R., and S. Hubbell, 1990 Abundance and DBA sequence of two-base repeat regions in tropical tree genome. Genome **34:** 66-71.

Dellaporta, S. L., J. Wood and J. B. Hicks, 1983 A plant DNA minipreparation: version II. Plant Molecular Biology Report **:** 19-21.

Do, N., and R. P. Adams, 1991 A simple technique for removing plant polysacchraide contaminants from DNA. Biotechniques **10:** 162-166.

Doyle, J. J., and J. L. Doyle, 1990 Isolation of plant DNA from fresh tissue. Focus **12:** 13-15.

Echt, C. S., and P. May-Marquardt, 1997 Survey of microsatellite DNA in pine. Genome **40:** 9-17.

El-Kassaby, D. G. W. E. a. Y. A., 1996a The biology and management of coniferous forest seeds: Genetic perspectives. The Forestry Chronicle **72:** 481-484.

El-Kassaby, Y. A., 1992 Domestication and genetic diversity - should we be conserned. Forestry Chronicle **68:** 687-700.

El-Kassaby, Y. A., 1996b Genetic variation in low elevation Douglas-fir of British Columbia and its relevance to gene conservation. Biodiversity and Conservation **5:** 779-794.

El-Kassaby, Y. A., and G. Namkoong, 1993 Impact of Forest Management Practices on Genetic Diversity and its Conservation, pp. 205-213 in *International Symposium on the Genetic Conservation and Production of Tropical Forest Tree Seed*, edited by R. M. Drysdale, S. E. T. John and A. C. Yapa, Saraburi, Thailand.

El-Kassaby, Y. A., and A. J. Thomson, 1990 Continued reliance on bulked seed orchard crops: is it reasonable?, pp. 56-65 in *Joint Meeting of Western Forest Genetics Association and IUFRO*, Olympia, Washington.

Fang, G., S. Hammar and R. Grumet, 1992 A quick and inexpensive method for removing polydaccharides from plant genomic DNA. Biotechniques **13:** 52-55.

Forests, B. M. o., 1993 Genetic Diversity in Forest Management, pp. . ministry of forests.

Fowells, H. A., 1965 Silvics of forest trees of the united states. Agricultural Hanbook **271:** 373-383.

Goodwin, D. C., and S. B. Lee, 1993 Microwave miniprep of total genomic DNA from fungi, plants, protists and animals for PCR. Biotechniques **15:** 1993.

Gross, D., and W. Garrard, 1986 The ubiquitous potential z-forming sequence of eukaryotes (dT-dG)n - (dC-dA)n is not detectable in the genomes of eubacteria, archaebacteria or mitochondria. Molecular and Cellular Biology **6:** 3010-3013.

Hamada, H., M. Seidman, B. Howard and C. Gorman, 1984 Enchanced gene expression by the poly(dT-dG) - poly(dC-dA) sequence. Molecular and Cellular Biology **4:** 2622-2630.

Hamrick, J., and M. Godt, 1990 Allozyme diversity in plante species, pp. 43-63 in *Plant Population Genetics, Breeding and Genetic Resources*, edited by A. Brown, M. Clegg, A. Kahler and B. Weir. Sinauer Associates, Sunderland, MA.

Hamrick, J. L., 1991 Allozyme diversity of natural stands verus seed orchard loblolly pine, pp. in *23rd biennial meeting of Canadian Tree Improvement Association*, Ottawa.

Hamrick, J. L., and M. J. W. Godt, 1989 Allozyme diversity in higher plants, pp. 53-67 in *Differentiation patterns in higher plants*, edited by K. Urbanska. Academic Press, New York.

Hong, Y.-K., D. A. Coury, M. Polne-Fuller and A. Gibor, 1992 Lithium chloride extraction of DNA from the seaweed *Porphyra perforata* (Rhodophyta). Journal of Phycology **28:** 717-720.

Hoopes, B. C., and W. R. McClure, 1981 Studies on the selectivity of DNA precipitation by spermine. Nucleic Acids Research **9:** 5493-5504.

Huff, D., R. Peakall and P. Smouse, 1993 RAPD variation within and among natural population of outcrossing buffalograss (*Buchloe dactyloides* (Nutt.) Engelm.). Theoretical and Applied Genetics **86:** 927-934.

Kawchuk, L., D. Lynch, J. Hachey, P. Bains and F. Kulcsar, 1994 Idnetification of a codominant polymorphic DNA marker linked to the verticillium wilt resistance gene in tomato. Theoretical and Applied Genetics **89:** 661-664.

Kenny, J. R., 1986 Nucleotide sequence of the carboxy-terminal portion of a lodgepole pine actin gene, pp. 100 in *Department of Forestry*. University of Alberta, Edmonton.

Khasa, P., and B. Dancik, 1996 Rapid indentification of white-Engelmann spruce species by RAPD markers. Theoretical and Applied Genetics **92:** 46-52.

Knowles, P., 1985 Comparison of isozyme variation among natural stands and plantations: jack pine and black spruce. Canadian Journal of Forest Research **15:** 902-908.

Ledig, F. T., 1992 Human impact on genetic diversity in forest ecosystems. Oikos **63:** 87-108.

Levinson, G., and G. Gutman, 1987 Slipped-Strand Mispairing: A major mechanism for DNA sequence evolution. Molecular Biological Evolution **4:** 203-221.

Liu, Z., and G. R. Furnier, 1993 Comparison of allozyme, RFLP, and RAPD markers for revealing genetic variation within and between trembling aspen and bigtooth aspen. Theoretical and Applied Genetics **87:** 97-105.

Lotan, J. E., 1975 The role of cone serotiny in lodgepole pine forests, pp. 86-104 in *Management of lodgepole pine ecosystems*, edited by D. M. Baumgarter, Pullman, Washington.

Maniatis, T., E. F. Fritsch and J. Sambrook, 1982 *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory, New York.

Marshall, D. R., and A. Brown, 1975 The charge state model of protein polymorphisms in natrual populations. Journal of Molecular Evolution **6:** 149-163.

Messmer, M. M., A. E. Melchieger, M. Lee, W. L. Woodman, E. A. Lee *et al.*, 1991 Genetic diversity among progenitors and elite line from the Iowa Stiff Stalk Synthetic (BSSS) maize populations: comparisons of allozyme and RFLP data. Theoretical and Applied Genetics **83:** 97-107.

Michelmore, R., I. Paran and R. Kessel, 1992 Indentifiaction of markers linked to disease resistance genes by bulked segregant analysis: A rapid method to detect markers in specific genomic regions using segregating populations. Proceedings from the National Academy of Sciences **88:** 9828-9832.

Millar, C. I., and K. A. Marshall, 1991 Allozyme variation of Port-Orford-cedar (*Chamaecyparis lawsoniana*): implications for genetic conservation. Forest Science **37:** 1060-1077.

Moran, G. F., and J. C. Bell, 1987 The origin and genetic diversity of *Pinus radiata* in Australia. Theoretical and Applied Genetics **73:** 616-622.

Morchem, M., J. Cuguen, G. Michaelis, C. Hanni and P. Saumitou-Laprade, 1996 Abundance and length polymorphism of microsatellite repeats in *Beta vulgaris* L. Theoretical and Applied Genetics **92:** 326-333.

Mosseler, A., K. N. Egger and G. A. Hughes, 1992 Low levels of genetic diversity in red pine confirmed by random amplified polymorphic DNA markers. Canadian Journal of Forest Research **22:** 1332-1337.

Mulcahy, D. L., 1975 Differential mortality among cohorts in a population of *Acer saccharwn* (Aceraceae) seedlings. American Journal of Botany **62:** 422-426.

Muller-Starck, G., and M. Ziehe, 1984 Reproductive systems in conifer seed orchards .Female and male fitnesses of individuals clones realized in seeds of *Pinus sylvestris* L. Theoretical and Apllied Genetics **69:** 173-177.

Muller-Strack, G., 1987 Genetic differentiation among seed samples from provenances of *Pinus sylvestris*. Silvae Genetica **36:** 232-238.

Murray, M. G., and W. F. Thompson, 1980 Rapid isolation of high molecular weight plant DNA. Nucleic Acids Research **8:** 4321-4325.

Naylor, L., and E. Clark, 1990 d(TG0n - d(CA)n sequences upstream of the rat prolactin gene from Z-DNA and inhibit gene transcription. Nucleic Acids Research **18:** 1595-1601.

Owens, J. N., and M. Molder, 1984 *The Reproductive Cycle of Lodgepole Pine*. Information Services Branch Province of B.C. Ministry of Forests, Vancouver.

Paran, I., R. Kesseli and R. Michelmore, 1991 Identification of RFLP and RAPD markers linked to downy mildew resistance genes in lettuce, using near-isogenic lines. Genome **34:** 1021-1027.

Pardue, M., K. Lowenhaupt, A. Rich and A. Nordheim, 1987 (dC-dA)n - (dG-dT)n sequences have evolutionarily conserved chromosomal locations in *Drosophila* with implications for roles in chromosome structure and function. EMBO **6:** 1781-1789.

Patel, M., 1992 The effects of promoter mutations on the developmental expression of dopa decarboxylase gene of Drosophila melanogaster, pp. 110 in *Department of Genetics*. University of Alberta, Edmonton.

Pfister, R. D., and R. Daubenmire, 1975 Ecology of lodgepole pine, pp. 27-46 in *Management of Lodgepole pine*, edited by D. M. Baumgartner, Pullman, Washington.

Piesch, R. F., 1987 Tree improvement comes of age in the Pacific Northwest - implications for the nursery manager. Tree Planter's Notes **38:** 3-8.

Rafalski, J. A., and S. V. Tingey, 1993 Genetic diagnostics in plant breeding: RAPDs, microsatellites and machines. Trend in Genetics **9:** 275-280.

Reiter, R., and J. Williams, 1992 Global and local genome mapping in *Arabidopsis thaliana* by using recombinant inbred lines and random amplified DNAs. Proceedings from the National Academy of Sciences **89:** 1477-1481.

Sarkar, G., and S. S. Sommer, 1990 Shedding light on PCR contamination. Nature **343:** 27.

Saunders, G. W., 1993 Gel purifiaction of red algal genomic DNA: an inexpensive and rapid method for the isolation of polymerase chain reaction-friendly DNA. Journal of Phycology **29:** 251-254.

Schlotterer, C., and D. Tautz, 1991 Slippage synthesis of simple sequence DNA. Nucleic Acids Research **20:** 211-215.

Skroch, P., and J. Nienhuis, 1995 Qualitative and quantitative characterization of RAPD variation among snap bean (*Phaseolus vulgaris*) genotypes. Theoretical and Applied Genetics **91:** 1078-1085.

Smith, D., and M. Devey, 1994 Occurrence and inheritance of microsatellites in *Pinus radiata*. Genome **37:** 977-983.

Stewart, C. N., and L. E. Via, 1993 A rapid CTAB DNA isolation technique for RAPD fingerprinting and other PCR application. Biotechniques **14:** 1993.

Taylor, B., and A. Powell, 1982 Isolation of plant DNA and RNA. Focus : 4-6.

Terauchi, R., and A. Konuma, 1994 Microsatellite polymorphism in *Dioscorea tokoro*, a wild yam species. Genome **37:** 794-801.

Tham, F., J. Lucas and Z. Wilson, 1994 DNA fingerprinting of *Peronospora parasitica*, a biotrophic fungal pathogen of crucifers. Theoretical and Applied Genetics **88:** 490-496.

Tingey, S., J. Rafalski and J. Williams, 1992 Genetic analysis with RAPD markers, pp. in *Joint Plant Breeding Symposia Series*. Crop Society of America

American Society for Horticultural Science

American Genetic Association, Minneapolis, Minnesota.

Tinker, D. B., W. H. Romme, W. W. Hargrove, R. H. Gardner and M. G. Turner, 1993 Landscape-scale heterogeneity in lodgepole pine serotiny. Canadian Journal of Forest Research **24:** 897-903.

Ven, W. V. D., and R. McNicol, 1995 The use of RAPD markers for the identification of Sitka spruce (*Picea sitchensis*). Heredity **75:** 126-132.

Weber, J. L., 1990 Informativeness of human (dC-dA)n (dG-dT)n polymorphisms. Genomics **7:** 524-530.

Weber, J. L., and P. E. May, 1989 Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. American Journal of Human Genetics **44:** 388-396.

Welsh, J., and M. McClelland, 1990 Fingerprinting genomes using PCR with arbitrary primers. Nucleic Acids Research **18:** 7213-7218.

Wilkie, S. E., P. G. Isaak and R. J. Slater, 1993 Random amplified polymorphic DNA (RAPD) markers for the genetic analysis in *Allium*. Theoretical and Applied Genetics **86:** 497-504.

Williams, J., M. Hanafey and S. Tingey, 1993 Genetic analysis using random amplified polymorphic DNA markers. Methods in Enzymology **218:** 704-740.

Williams, J. G. K., and et al., 1990 DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucleic Acids Research **18:** 6531-6535.

Wu, K., and S. Tanksley, 1993 Abundance, polymorphism and genetic mapping of microsatellites in rice. Molecular and General Genetics **241:** 225-235.

Appendix A

This is an example of the procedure used to generate chi-square values to determine if the difference in allele frequency estimates from each treatment or population differ significantly.

The example used is from a comparison between unharvested and planted treatments at locus one. The values for q^2, the frequency of the homozygous recessive genotype, obtained from the RAPD gels are outlined in the table below.

Treatment	q2	Total individuals
Unharvested	0.17	90
Planted	0.18	89

Once the values of q^2 have been obtained, we can calculate the allele frequency by taking the square of the q^2 shown below.

Treatment	q	1-q	Total alleles
Unharvested	0.41	0.59	180
Planted	0.42	0.58	178

Note: that the total alleles is twice the number of total individuals since lodgepole pine is 2n. We can now calculate the observed number of alleles present by multiplying the frequency by the total.

Treatment	q	1-q	Total
Unharvested	73.48	106.51	180
Planted	75.47	102.53	178
Total	148.95	209.04	358

Once all of the observed values have been calculated, an expected value can then be calculated for each treatment. For example, the expected q for unharvested would be calculated by multiplying 148.56 by 180 and dividing by 358 giving you a value of 74.69. The expected values for each of the treatments are displayed below.

Treatment	q	1-q	Total
Unharvested	74.89	105.11	180
Planted	74.06	103.94	178
Total	148.95	209.05	358

Now a simple 2x2 chi-square analysis can be conducted to determine if the differences are significant. The chi-square value obtained is 0.09 which is not significant at the 5% level.